MACHINE LEARNING FOR CORPORATE FAILURE PREDICTION – AN EMPIRICAL STUDY OF SOUTH AFRICAN COMPANIES

Prepared by:
Saul Kornik

Prepared for:
The Department of Accounting and the Department of Electrical Engineering at the University of Cape Town

31 March 2004

Dissertation prepared in fulfilment of the requirements for the Degree of Masters in Commerce

ACKNOWLEDGEMENTS

The research performed in this report would not have been possible without the time and effort of Professor Geoff Everingham and Professor John Greene. Professor Everingham, of the Commerce Faculty at the University of Cape Town, provided an invaluable contribution in the guiding of the financial aspects of the construction of the corporate failure prediction model. Professor Greene, of the Electrical Engineering Department at the University of Cape Town, was tireless in coaching me in the concepts behind the machine learning techniques used in this study. In addition, his insight and understanding of these methods was a vital resource at times of critical decision-taking.

I would also like to take this opportunity to thank Lars Piepke, a director of Decillion Fund Management (Pty) Ltd, for his assistance in the preliminary analysis of the sample of companies used in this study.

SYNOPSIS

The research objective of this study was to construct an empirical model for the prediction of corporate failure in South Africa through the application of machine learning techniques using information generally available to investors.

The study began with a thorough review of the corporate failure literature, breaking the process of prediction model construction into the following steps:

- Defining corporate failure
- Sample selection
- Feature selection
- Data pre-processing
- Feature Subset Selection
- Classifier construction
- Model evaluation

These steps were applied to the construction of a model, using a sample of failed companies that were listed on the JSE Securities Exchange between 1 January 1996 and 30 June 2003. A paired sample of non-failed companies was selected. Pairing was performed on the basis of year of failure, industry and asset size (total assets per the company financial statements excluding intangible assets). A minimum of two years and a maximum of three years of financial data were collated for each company. Such data was mainly sourced from BFA McGregor RAID Station, although the BFA McGregor Handbook and JSE Handbook were also consulted for certain data items.

A total of 75 financial and non-financial ratios were calculated for each year of data collected for every company in the final sample. Two databases of ratios were created – one for all companies with at least two years of data and another for those companies with three years of data. Missing and undefined data items were rectified before all the ratios were normalised.

The set of normalised values was then imported into Matlab Version 6 and input into a Population-Based Incremental Learning (PBIL) algorithm. PBIL was then used to identify those subsets of features that best separated the failed and non-failed data clusters for a one, two and three year forward forecast period. Thornton's Separability Index (SI) was used to evaluate the degree of separation achieved by each feature subset.

After running the PBIL algorithm numerous times, it was discovered that a number of feature subsets, each composed of significantly different features but still with near-identical separability, could be identified.

The optimal feature subsets for each forecast period were then selected and all other feature subsets were discarded. Only one feature subset was identified as optimal for the purposes of the one year forward forecast model. In the case of the two year forward forecast model, two different feature subsets with equal SI values were identified. Three different feature subsets were brought forward into the classifier construction stage for the three year forward forecast model.

A k-Nearest Neighbour (kNN) and Kernel Ridge Regression (KRR) classifier were then run on each of the optimal feature subsets. The parameters for each model were determined using heuristic procedures.

- A value of "1" for k in the kNN classifier was determined as optimal.
- The value of gamma (the regularisation parameter) in the KRR classifier was of no consequence to the accuracy of this classifier. This was interpreted as evidence that the data sets were well-conditioned. It was also held as an indication that the data clusters were overlapping rather than distinctly separate with occasional outliers.
- The value of sigma (kernel width), however, was determined to be a critical parameter in the implementation of the KRR classifier.

In this way, a 1-Yr forward forecast model, two 2-Yr forward forecast models and three 3-Yr forward forecast models were constructed. Each classifier was trained and then evaluated using the Leave-One-Out (LOO) validation method. The predicted company classifications relating to each model were collated and the type I and type II error rates were calculated.

The relative error rates were used to evaluate the comparative performance of the models for each forecast period. In addition, misclassification costs were calculated for each model using 20:1, 1:1, and 1:20 cost ratio assumptions (type I: type II). The relative performance of these models was then evaluated using these costs.

There did not appear to be a single classification technique or feature subset that outperformed all others on all accounts. The performance of the different feature subsets varied with the application of different cost assumptions. Furthermore, KRR outperformed kNN when those subsets that contained a larger number of features were input into the classifier. Conversely, kNN outperformed KRR with the smaller feature subsets.

Throughout the study, numerous topics for further research, that did not fall within the scope of this study, were identified. These are presented at the end of this report.

Over and above presenting the final models that were constructed, this report seeks to justify the procedures and elections made at each stage in the corporate failure prediction model construction process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1. BACKGROUND TO RESEARCH: AN INTRODUCTION TO CORPORATE FAILURE

Corporate failure is an essential component of an efficient market economy. It allows for the recycling of financial, human and physical resources into more productive organisations (Easterbrook, 1990; Schumpeter, 1934). However, many stakeholders, including shareholders, providers of debt finance, employees, suppliers, customers, managers and auditors, have an interest in the financial health of a firm, as the failure of the corporation can have a significant impact on the costs to all of these parties.

Such costs can be reduced if the trajectory towards corporate failure is identified early. As a result, there has been extensive research over the past 40 years into developing failure forecasting models. Some studies have attempted to forecast such failure as long as ten years in advance of the ultimate collapse (Hambrick & D'Aveni, 1988).

1.1.1. Defining Corporate Failure

Corporate failure can arise for many reasons. It may occur due to a single catastrophic event or it may be the end result of a lengthy process of decline (Brabazon et al, 2001). Under the latter scenario, corporate failure can be seen as a process which may start with management defects, leading to poor decisions, leading to financial deterioration and finally leading to corporate collapse (Altman, 1993). Most attempts to predict corporate failure implicitly assume that management decisions critically impact on firm performance (Argenti, 1976).

The premise of this report is that a series of poor financial decisions leads to the deterioration in the financial health of the firm and finally to its demise. Although the decisions are not directly observable, their consequent affects on the financial health of the firm can be observed.

As the accurate prediction of a firm's failure is of little cost advantage at the point of its demise, it is more beneficial to construct a model that can predict when a firm has reached an earlier stage in the process of decline. There is little consensus in the literature as to the level of financial ill-health at which it is optimal to define failure.

1.1.2. MEASURING THE SYMPTOMS OF CORPORATE FAILURE

Studies in the area of corporate failure have utilised a number of different explanatory variables in order to observe the consequences of poor financial decisions having been made. Such variables have included company financial statement information, general macro-economic data and non-financial firm-specific data (discussed in Section A).

Due to the massive quantity and complicity of available information, a key step in the construction of a failure prediction model is to distil the available data so that only the relevant data remains for input into the prediction process.

1.1.3. TRADITIONAL CORPORATE FAILURE PREDICTION METHODS

Beaver (1966) was the first to employ traditional statistical methods for corporate failure prediction. He used a univariate methodology to identify which accounting ratios had the greatest classification accuracy when identifying failing and non-failing firms. Subsequent to this study, various multivariate models have been employed by researchers in order to assess several financial facets of a corporation simultaneously. Discriminant, Probit and Logit analysis have been the most commonly applied multivariate methodologies per the literature. The history of the application of traditional statistics to the problem of corporate failure is discussed in detail in Section A.

More recently, the methodologies applied to this problem have included neural networks, genetic algorithms and other advances in machine learning techniques.

1.2. POTENTIAL FOR THE APPLICATION OF MACHINE LEARNING METHODOLOGY TO THE PREDICTION OF CORPORATE FAILURE

Machine learning techniques are applied in research domains ranging from the diagnosis of medical test data to plant disease. Machine learning can broadly be defined as the field of study that concentrates on algorithms that have the ability to learn. Learning is an even broader term which denotes the gaining of knowledge, skill and understanding from instruction, experience or reflection (Easterbrook, 1990, 411). This is in direct contrast to expert systems that are automated with a set of predetermined rules for the classification of the independent variable.

Machine learning techniques are adept at finding potential solutions to highly complex problems. It is imperative in the application of such techniques, that there does exist a well-defined underlying relationship between the explanatory variables and the classes to be predicted.

There are two facets to corporate failure prediction model construction to which machine learning techniques may successfully be applied: explanatory variable subset selection and the development of a classifying function (each discussed further below). This report applies Population-Based Incremental Learning (PBIL) to the former and k-Nearest Neighbour (kNN) and Kernel Ridge Regression (KRR) to the latter task.

1.2.1. EXPLANATORY VARIABLE SUBSET SELECTION

A multitude of data is available from company financial statements. While there has been much documented on the way in which to combine various pieces financial statement information in order to assess the financial position and performance of a company, there has been little consensus on which subsets best predict corporate failure.

The selection of the subset of explanatory variables that best defines the difference between failed and non-failed firms represents a high-dimensional combinatorial problem well suited to a stochastic optimisation algorithm. The possibility that there may be numerous optimal solution subsets was a consideration factored into the application of the Population-Based Incremental Learning (PBIL) algorithm applied in this study.

1.2.2. DEVELOPMENT OF A CLASSIFYING FUNCTION

Inductive machine learning uses specific examples to draw general conclusions in identifying relationships within a set of data (i.e. learning by examples).

Classical statistical methods are capable of dealing with such problems. However, the validity and effectiveness of such methods is largely dependent on restrictive assumptions such as linearity, normality, independence and homoskedacity (discussed in detail in Chapter 8).

The inductive learning process of KRR does not rely on such distributional and statistical assumptions in deriving its model coefficients. Rather, the iterative process searches heuristically for the coefficient set that minimises a defined loss function. However, certain parameters still need to be user-selected. Such optimal parameter values can also be isolated through a heuristic procedure.

Likewise, the parameter selection process for the kNN algorithm does not rely on restrictive distributional and statistical assumptions.

1.3. DEFINING THE RESEARCH PROBLEM

A number of different studies have applied machine learning techniques to the problem of corporate failure prediction (see Chapter 9.4.). While certain studies achieved superior results using such methods, there has been no conclusive evidence indicating that inductive classifiers outperform conventional statistical methods in this area of research.

However, machine learning is a relatively new concept in comparison to the conventional statistical methods that have saturated corporate failure prediction since the birth of this empirical field of study in the 1960's. In addition, there have been many advances made and much inter-disciplinary research performed in the area of machine learning.

In South Africa, while multi-layer perceptron neural networks have been applied to this prediction problem, there has been little research using the more recently developed machine learning techniques.

With this in mind, the research objective of this study was defined as follows:

> *This study seeks to construct an empirical model for the prediction of corporate failure in South Africa through the application of machine learning techniques using information generally available to investors.*

1.4. SCOPE AND LIMITATIONS OF RESEARCH

This empirical study considered those companies that were listed on the JSE Securities Exchange and subsequently failed within the period 1 January 1996 to 30 June 2001. The report presents a model constructed to predict such failure one, two and three financial year ends prior to the date of failure using firm-specific financial and non-financial data.

The scope of this study was limited to two of the major paradigms of machine learning:

- inductive learning (or learning by example); and

- stochatistic (or evolutionary) optimisation algorithms, i.e. randomised methods that search for a solution without making explicit structural descriptions of the search space.

At each stage of the research and construction process, the available run-time resources were traded off against the value of the additional information that could be obtained. This restricted the scope of the study to one that was manageable based on the resources that were available.

In addition, this study did not seek to make any direct comparisons of its results to those of other similar studies. Direct comparison is potentially misleading because of the different time periods, geographic locations, assumptions and data availability under which different studies have constructed their prediction models. However, each step of the process still draws on the published research as it relates to the construction of the model in this study.

Finally, a note should be made that the base assumptions underlying the use of data in this study were that:

- financial markets are efficient; and
- creative accounting tactics have not been employed by management to manipulate the financial statement information.

1.5. STRUCTURE OF REPORT

The development of corporate failure prediction models has been a thoroughly researched topic in the area of finance. Corporate failure studies follow a number of common and necessary steps in the construction of such models. These steps are summarised in the table below.

This report is divided into five sections:

- Section A provides a thorough discussion of the literature surrounding each of the steps involved in corporate failure prediction model construction.
- Sections B to D contain the research performed in the construction of the prediction model in this study. These sections draw on the research reviewed in Section A while providing a review of literature related to the machine learning techniques applied in this study.
- Section E concludes and presents areas for further research.

The table below sets out which chapters in both Section A and Sections B to D deal with which steps of model construction.

Step in Model Construction	**Section A: Review of the Literature**	**Sections B to D: Construction of Model**
Corporate Failure Definition	Chapter 3	Chapter 13
Failed and Non-Failed Company Sample Selection	Chapter 4	Chapter 14
Feature Set Selection	Chapter 5	Chapter 15
Data Collection	Chapter 6	Chapter 15
Data Pre-Processing	Chapter 6	Chapter 16
Feature Subset Selection	Chapter 7	Chapter 17 & 18
Classifier Construction	Chapter 8 & 9	Chapter 19 & 20
Model Evaluation	Chapter 10	Chapter 21

Table 1.1. Structure of this report

SECTION A
LITERATURE REVIEW OF EMPIRICAL RESEARCH ON CORPORATE FAILURE

CHAPTER 2
INTRODUCTION: REVIEW OF LITERATURE

Since the construction of business failure prediction models became a field of study, researchers have introduced a plethora of methods to address the different facets of such models. These methods differ according to the views and requirements of the researcher. Each new study "tweaks" a facet in a manner that has not been addressed in the preceding research – some successfully and others unsuccessfully.

There are also numerous published articles that exhaustively summarise the diversity and scope of this subject of research. These include Scott (1981), Altman (1984), Dimitritas, Zanakis & Zopounidis (1996) and Atiya (2001), as reviewed in this section.

2.1. COMMON STEPS IN CORPORATE FAILURE PREDICTION MODEL CONSTRUCTION

As noted in Chapter One, there are a number of common steps involved in the process of building any model for corporate failure prediction. Studies have differed in the emphasis placed on the various stages of this process.

The literature, as reviewed in this section, is organised into these steps, briefly summarised as follows:

- **Corporate Failure Definition:**
  There has been little consistency in the literature regarding the definition of this concept. This definition is critical to such research for two main reasons. Firstly, this definition will determine which of the sampled companies are labelled as failed or non-failed, respectively. Also, as the ultimate goal of model construction is to be able to predict the state of a company prior to its failure, the definition will determine what state such model attempts to predict. The various definitions of corporate failure, as applied in the literature, have been reviewed in Chapter Three.

- **Failed and Non-Failed Company Sample Selection (Dependent Variable):**
  The corporate failure definition is then applied in the company sampling procedure in order to distinguish between failed and non-failed firms. The vast majority of corporate failure studies, including this one, employ this failed/non-failed

dichotomous classification. Various sample selection procedures are discussed in Chapter Four.

- **Feature Set Selection (Independent/Predictor Variable):**
  This step involves selecting the entire set of features from which the significant predictor subset will be chosen. These variables can be firm-specific or macroeconomic in nature. The accuracy of the model is dependent on a relationship existing between these predictor variables and the classification of the sample of companies as failed or non-failed. Predictor variable selection is reviewed in Chapter Five.

- **Data Collection:**
  All the data required to calculate the set of features selected in the previous step, are then collected for each company in the sample. Chapter Six discusses this in further detail.

- **Data Pre-Processing:**
  The predictor variable data is analysed and transformed so as to maximise the distinction between the failed and non-failed companies. Various pre-processing techniques are discussed in Chapter Six.

- **Feature Subset Selection:**
  The set of features is then tested in order to determine which subset will best explain the distinction between the classification of a company as failed and non-failed. Different methods for selecting this optimal subset of predictors are discussed in Chapter Seven.

- **Classifier Construction:**
  Many different techniques have been applied to this research problem since the inception of empirical research in this area in the 1960's. The literature includes techniques drawn from medical, engineering and other spheres of research. Each method has its own advantages, limitations and assumptions. Chapter Eight reviews the more commonly applied techniques. Chapter Nine discusses the history of machine learning as applied to this field of research.

- **Corporate Failure Prediction Model Evaluation:**
  The final evaluation of the model(s) constructed involves testing either the model itself or the classification outcomes for significance. Various evaluation techniques are outlined in Chapter Ten.

2.2. BRIEF HISTORICAL REVIEW OF FAILURE PREDICTION MODEL DEVELOPMENT

2.2.1. PRIOR TO BEAVER (1966)

Prior to the development of quantitative measures of company performance, agencies were established to supply qualitative information that could be used to assess the creditworthiness of a particular business. For instance, the "forerunner" of Dun & Bradstreet, Inc. was organised in 1849 in Cincinnati, Ohio, in order to provide independent credit investigations (Altman, 1968, 590).

Altman, in his seminal paper, noted that a study by R.F. Smith and A.H. Winakor titled "Changes in the Financial Structure of Unsuccessful Corporations" (University of Illinois: Bureau of Business Research, 1935), and several later studies, concluded that failing firms exhibit significantly different financial ratio measurements than surviving entities. These studies, however, were not able to quantify the relationship between financial ratios and corporate failure.

Only since studies by Beaver (1966) and Altman (1968), has any serious attempt been made to quantify the prediction of corporate failure. Sharma & Mahajan (1980, 80) suggest the following reasons for the prior lack of interest in this area of study:

- the negative connotation of the term "failure";
- the notion that the failure process for a firm is atypical and, hence, does not lend itself to a scientific study;
- the lack of an available and published body of knowledge relating to failure;
- the belief that failure is a sudden rather than a gradual process.

2.2.2. EMPIRICAL PREDICTION MODEL RESEARCH

The prominent developments in empirical models for the prediction of corporate failure have been summarised into the time line below. This is intended to provide an overview of such developments and should not be considered, by any means, exhaustive.

1966 Beaver:

Beaver used a **univariate analysis** technique to determine the effectiveness of **individual financial ratios** in discriminating between failed and non-failed firms. His definition of failure was **broader** than simple bankruptcy.

1968 Altman:

The Z-Score: Reconfirmed the usefulness of **financial ratio analysis** and achieved improved predictive power through the use of a multiple variable approach - **multivariate discriminant analysis (MDA)**. The definition of failure was narrowed to **bankruptcy**. However, Altman did not develop a theoretical framework to back his findings and he failed to avoid some of the MDA statistical pitfalls.

1972 Deakin:

The accuracy of Altman's model was improved by using **all fourteen ratios** that Beaver had identified as good predictors of bankruptcy. Deakin used a **probabilistic classification rule,** rather than a subjective cut-off point, to determine the critical discriminant Z-score at which to classify a firm as having failed. He made statistical errors similar to those made by Altman (1968).

1974 Blum:

In this study, the **Failing Company Doctrine** was used to define failure. Blum used a **linear MDA** model and included financial variables expressed in terms of **change over time**. He stressed that it is imperative that a failure prediction model have **theoretical justification** and soundly **validated procedures**. This study has some statistical shortcomings.

1975 Libby:

Libby was the first to use **principal component (factor) analysis** to reduce the dimensionality of a data set and select only those financial variables that add new information. He investigated the **behavioural aspects** of failure prediction by testing the ability of loan officers to evaluate ratio information.

1977 Altman, Haldeman & Narayanan:

ZETA model: Explicitly considered **developments** in the business failure field at the time of the paper. A more sophisticated and **statistically correct MDA** was employed. The model explicitly considered **costs of misclassification** and ***a priori* probabilities**.

1980 — Ohlson:

Ohlson introduced the **logit model** to bankruptcy prediction. This model differed in that a probability value was attached to each failure prediction. He was the first to use a **non-paired sample technique** to select non-failed companies. The data set used included **size** as a predictor variable for the first time.

1984 — Whittred & Zimmer:

Whittred and Zimmer, in an Australian study, included the non-financial variable of **reporting time delay** in their model in order to improve its predictive ability.

1985 — Frydman, Altman & Kao:

This study introduced **recursive partitioning analysis (RPA)** as a classification procedure. RPA is nonparametric in nature. The statistical assumption criticisms of parametric techniques were avoided by using this method.

1986 — Lane, Looney & Wansley:

Lane et al, introduced the use of the **Cox (1972) proportional hazards** method of classifying banks as failing. The Cox model calculates an estimated **time to failure**.

1986 — Peel, Peel & Pope:

This study calculated empirical results for the U.K., incorporating **firm-specific non-financial variables** as extracted from the company annual reports.

1987 — Lau:

Lau moved away from the dichotomous classification of firms as failed and non-failed. She constructed a model that calculated the probability of a firm falling into **one of five states on a continuum** of corporate financial health.

1994 — Platt, Platt & Pederson:

Platt et al, considered **macroeconomic factors** together with firm-specific financial ratios in their failure prediction model construction. Both **deflators** of nominal financial ratios and additional relevant **macroeconomic variables** were included in alternately constructed models.

The above diagram chronologically tabulates some of the key points in the development of corporate failure prediction models. It does not include a review of those models constructed using machine learning techniques. For a summary of the development of neural network models as applied to this field of research, refer to Chapter 9.4.2.

The above time line illustrates the progression in the types of classification techniques used (from univariate to multivariate to various non-parametric methods), information employed (firm-specific financial and non-financial, as well as macroeconomic), and the manner in which the corporate failure concept has been defined. It also serves as a basic illustration of the ongoing debate and search for improvement that characterises this field of study.

CHAPTER 3
DEFINITION OF CORPORATE FAILURE

Although the idea of a company going bankrupt is associated with its disappearance, the company actually goes through a period of crisis, consisting of several stages, before such a demise. Some studies have justified defining the corporate failure concept at one of these earlier stages. This is a critical decision in the process of model construction.

3.1. THE CORPORATE FAILURE SPECTRUM

> "**Failure**: Lack of success, non-performance, breaking down or ceasing to function." (Oxford Dictionary, 1998)

If failure is defined as expressed above, then the concept of corporate failure embodies a broad range of scenarios less critical than bankruptcy. While "ceasing to function" captures failure in law, "lack of success" implies failure as measured against some economic criteria representing "success". Van Horne (1986, 741) found the term confusing. He pointed out that "the word failure is vague, partly because there are varying degrees of failure".

As noted by Court (1983, 7), these "varying degrees of failure" are classified in the literature on a general spectrum ranging between economic failure, insolvency, financial failure and bankruptcy. This spectrum can be summarised as follows:

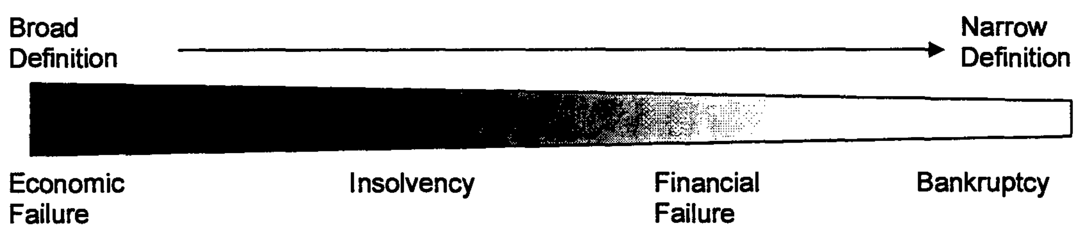


Figure 3.1. Spectrum of Corporate Failure

3.1.1. Economic Failure

Altman (1983, 6) defined failure as the situation in which "the realised rate of return on invested capital, with allowances for risk considerations, is significantly and continually lower than prevailing rates on similar investments". Platt (1985, 7) defined the term even more broadly as "when the business is not sufficiently prosperous given the level of capital investment and human effort put into making it work". These definitions seek to define the corporate failure concept in terms of a broader meaning which encompasses the needs of a rational investor.

3.1.2. Insolvency

Insolvency is more of a technical term. Altman (1983, 6) defined insolvency as: a situation in which a company has insufficient cash inflows to meet its current obligations. In this position, Dimitras et al (1996, 487) suggested that the firm has two options other than to cease operations - to liquidate (discussed further later in this chapter) or to reorganise the business into a more liquid venture.

3.1.3. Financial Failure

Court (1983, 8) and Van Horne (1986, 741) described financial failure as covering the spectrum between the weak insolvency position and the final bankruptcy declaration.

3.1.4. Bankruptcy

Altman (1983, 7) defined two types of bankruptcy. The one type is a **technical insolvency** where the liabilities are greater than the fair value of the assets of the company. The real net worth of the company is negative. The second is a **formal declaration of bankruptcy in court** that would usually follow the technical insolvency of a company.

According to this spectrum, bankruptcy is included as a narrow part of the definition of economic failure, insolvency and financial failure.

3.2. DEFINING CORPORATE FAILURE IN THE LITERATURE

3.2.1. Possible Explanations for Variations in Applied Definitions

A single unifying definition of failure has not been applied consistently across the literature. The definition, as applied in different studies, has depended on what concept of failure the researcher is attempting to predict. The motivation for constructing models that attempt to predict a certain concept of failure will depend at which stakeholder in the firm such a model is targeted.

There are, however, other reasons of a practical nature that may impact on the selection of a definition for corporate failure:

- The process of legal failure and legal reorganisation differs across countries.
- The scope and availability of historical data on failed companies may be limited. In addition, the costs involved in obtaining such data will vary across geographic location and time period. As a result, the definition for the selection of failed companies may be tailored to suit that information that is readily available.
- There are different data requirements for the various methodologies applied to the problem of corporate failure prediction. In certain circumstances, the researcher may need to broaden the definition in order to allow for an enlarged sample of failed companies. In this way the sample size requirements of the selected methodology can be met.

Once failure has been defined, the events that represent the symptoms of the specific failure definition selected for a particular study need to be identified. These events are then used to apply the definition to the failed company selection process.

3.2.2. CLASSICAL CORPORATE FAILURE STUDY DEFINITIONS

The following are a few examples of some of the common definitions and application criteria used for the selection of failed companies in key studies outside of South Africa.

(a) Beaver (1966, 71)

Beaver, a seminal researcher in corporate failure prediction, defined failure as "the inability of a firm to meet its financial obligations as they mature". Operationally, a firm was said to have failed if any of the following events had occurred:

- Bankruptcy
- Bond default
- Non-payment of preference share dividend
- Overdrawn bank account

As noted by Altman, Avery, Eisenbeis & Sinkey (1981), these four events are extremely heterogeneous. An overdrawn bank account is far less serious than filing for bankruptcy. Similarly, bond default and non-payment of preference dividends are more serious than an overdrawn bank account, yet not as final as bankruptcy. In defining corporate failure in this manner, Beaver considered a broad range of the corporate failure spectrum within the scope of his study.

(b) *Altman (1968, 593)*

Altman, the pioneer of modern corporate failure prediction, defined corporate failure in the formal legal sense. Companies were included in his sample if they had "filed a bankruptcy petition under Chapter X of the National Bankruptcy Act".

Most U.S. studies have used a similar narrow definition. This is possible in the U.S. because of the formality of the bankruptcy petition, the number of companies that have filed for bankruptcy and the database of information available on such companies.

(c) *Blum (1974, 3)*

Blum looked to the legal precedent of the International Shoe Failing Company Doctrine (*International Shoe v. F.T.C., 280 U.S. 291 (1930)*). The Doctrine was used in merger law in order to define the point of failure of a company. According to the case, one of the following three events constitutes failure:

- the inability to pay debts as they become due,
- the entrance into bankruptcy proceedings, or
- an explicit agreement with creditors to reduce risk.

Ninety percent of Blum's selected failed firms filed bankruptcy petitions.

(d) *Other studies*

Several studies have considered other alternatives to the finality of bankruptcy. These include:

- liquidation (Truter, 1996), discussed later in this chapter;
- reorganisation (Casey et al, 1986);
- acquisition by, or merger with, a healthy company (Bulow & Shoven, 1978; Altman et al, 1977); and
- in the study of bank failures, poor supervisory ratings have been used (Korobow & Stuhr, 1984).

3.3. DEFINITIONS FOR FAILED COMPANY SELECTION IN SOUTH AFRICAN STUDIES

Studies published in South Africa have historically defined the corporate failure concept in terms of a broader economic meaning, as defined on the spectrum illustrated in Figure 3.1. This trend is clearly illustrated in the studies reviewed below.

There are a number of practical reasons for such a trend:

- There is a small population of failed companies in South Africa, relative to many other countries, from which the sample can be drawn.
- The availability and access to the financial information of such companies is limited.

For these reasons, South African studies have tended to define the failure concept broadly so that there is a sufficiently large data set and population of companies from which to draw the sample.

In addition, it is not possible simply to include those companies that have been liquidated in South Africa in the failed company sample - as one can do with bankrupt companies in the U.S. This is because the formal liquidation process in South Africa may be instituted for a number of non-failure related reasons. This is discussed further below.

The table below summarises the different criteria for failed company selection in the reviewed South African corporate failure studies. A more detailed discussion follows.

	Criteria for Failed Company Selection.	South African Studies
(1)	Company liquidated	Daya (1977) Immelman (1980) Le Roux (1980) De la Rey (1981) Walters (1982) Merks (1986) Arron (1994) Truter (1996)
(2)	Company placed under judicial management	Immelman (1980) Le Roux (1980) De la Rey (1981)
(3)	Company had a negative net worth	Immelman (1980) De la Rey (1981)
(4)	Company reduced share capital to bring it in line with related assets	Immelman (1980) De la Rey (1981)
(5)	Company failed to honour debt commitments	De la Rey (1981) Merks (1986)
(6)	Company failed to pay preference dividends	Immelman (1980) De la Rey (1981)
(7)	Company had poor financial performance	Daya (1977) Immelman (1980) De la Rey (1981) Court (1983)
(8)	Company failed to pay ordinary dividend	Immelman (1980) Le Roux (1980) De la Rey (1981)
(9)	Takeover of a company coupled with other criteria	Immelman (1980) Le Roux (1980)

Table 3.1. Criteria for failed company selection as utilised in South African corporate failure studies

3.3.1. COMPANY LIQUIDATED

(a) The law of liquidation in South Africa:

The process of liquidation or winding-up in South Africa is defined in *The Laws of South Africa* (Blackman, 1996) as when:

> "...the management of a company's affairs is taken out of its directors' hands, its assets are ascertained, realised and applied in payment of its creditors according to their order of preference, and any residue distributed among its members according to their rights. The company's corporate existence is then put to an end by the formal process of dissolution." (para. 98)

In South Africa, the Companies' Act No. 61 of 1973 provides for two modes of winding-up - voluntary winding-up and compulsory winding-up by the court.

A **voluntary winding-up** is initiated by a special resolution of the members of a company. A company may enter into such a voluntary liquidation process for a number of reasons that the members may consider to be worthwhile. These may range from near insolvency and poor financial performance (failure related reasons) to a company having reached the end of its economic life, as with a depleted mining operation.

A **winding-up by the court** (or **compulsory winding-up**) may be instituted on a number of different grounds provided for in Section 344 of the Companies' Act. A company can be wound-up by the court if it has lost seventy-five percent of its issued share capital (s. 344 (e)) or if it is unable to pay its debts (s. 344 (f)). However, non-failure related reasons, such as irregularities in the commencement of business and the reduction of the number of directors of a public company to below seven, are also grounds for the liquidation of a company by the court. Thus, the compulsory liquidation of a South African company may not necessarily indicate the failure of such a company.

In order to wind-up a company that is unable to pay its debts (section 344 (f)), it must be proved that the company is unable to pay such debts in terms of Section 345. In the case *ABSA Bank Ltd v. Rhebokskloof (Pty) Ltd 1993 4 SA 436 (C)*, the court set a precedent for proving this:

> "A company is in fact unable to pay its debts when it is unable to meet current demands on it, or its day-to-day liabilities in the ordinary course of business; in other words, when it is "commercially insolvent"." (Blackman, 1996, paragraph 123)

Hence, a company can be commercially insolvent (indicating a cash flow problem) and yet still be factually or technically solvent (the fair value of its assets exceeds all liabilities owing by the company), and so be subject to compulsory liquidation. Factual insolvency alone is not grounds for the winding-up of a company under Section 344 (f).

(b) Liquidation as a criterion for failed company selection in South African literature

From the discussion above, it is clear that viewing all liquidations of companies as an indication of failure may be incorrect. Immelman (1980) and Le Roux (1980) did not distinguish between the possible reasons for liquidation and seem to have included all companies that have been liquidated in their sample of failed companies.

The other studies qualified this "liquidation criteria" by referring specifically to companies that had gone insolvent or had been wound-up for other failure related reasons. Walters (1982) referred specifically to companies wound-up by the court when the company was deemed unable to pay its debts according to Section 345 of the Companies' Act.

3.3.2. COMPANY PLACED UNDER JUDICIAL MANAGEMENT

(a) The law relating to judicial management in South Africa

Judicial management is governed by Sections 427 to 440 of the Companies' Act, No.61 of 1973. A company is placed under judicial management by the court when (s.427):

- it is unable to pay its debts and is prevented from becoming a successful concern for some reason that could possibly be resolved through judicial management, or
- the same parties that can apply to have a company wound-up, apply for the judicial management of that company, or
- there has been an application to wind-up a company by the court and the court decides that the company should rather be placed under judicial management.

The judicial manager, appointed by the court, then attempts to bring the company back to solvency. If successful, the judicial management order will be lifted and the company will return to ordinary business. Failure of judicial management may result in the company being placed into final liquidation.

(b) Judicial management as a criterion for failed company selection in South African literature

Companies that are placed under judicial management have failed in some facet of their operations and are provisionally being rescued. Similarly, companies placed under provisional liquidation may in the interim attempt to reorganise their business through a Section 311 scheme of arrangement. Proponents of this failure criterion have argued that both judicial management and the scheme of arrangement indicate that the business has already failed and that measures of a last resort are being taken in order to salvage the business.

3.3.3. COMPANY HAS A NEGATIVE NET WORTH

A company is technically or factually insolvent when its liabilities (including future and contingent liabilities) exceed the fair value of its assets (LAWSA, paragraph. 112). As fair value and contingent liability information has, historically, not been readily available from the financial statements of a company, historic cost accounting values and existing recognised liabilities have been used instead.

Once a company reaches such a state, many researchers have considered the company to have declined to such an extent that drastic measures are being instituted in order to prevent its ultimate demise. As such, the company is considered to have failed.

3.3.4. COMPANY REDUCED SHARE CAPITAL TO BRING IT IN LINE WITH RELATED ASSETS

Immelman (1980, 5) noteed that where the financial position of a company is weak and where there are negative reserves, some of the share capital has effectively been lost. In certain situations the company may write these losses off against share capital. Until 2001, the reduction of share capital in South African companies was only permitted after application to the court.

As such, a reduction of this nature was considered to be a measure of last resort, largely because of the signal that it sent to investors. For this reason, this action was considered to embody a measure of financial failure.

3.3.5. COMPANY FAILED TO HONOUR DEBT COMMITMENTS

De la Rey (1981), in defining his criteria for failed company selection, made reference to companies that had not been able to pay their debts in a timely manner. Merks (1986) made reference to companies that had a cash flow problem resulting in the non-payment of debt. Such default might only be discovered after the company is placed into liquidation. However, with access to the databases of banks and other providers of finance, late payment and default can be identified.

Such a default results in an increase in the cost of capital to the company. Caused by a weak cash flow position, such a scenario was considered to be an indication of financial failure.

3.3.6. COMPANY FAILED TO PAY PREFERENCE DIVIDENDS

De la Rey (1981) considered any late payment of preference dividend to be indicative of failure.

Although there is no contractual commitment on the part of the issuer to pay preference dividends, failure to pay this dividend may be symptomatic of insufficient cash reserves or poor financial performance, implying a form of a corporate failure. However, there may be situations in which a company may choose not to pay the dividend in order to invest the cash in available profitable projects. De la Rey deemed this unlikely, arguing that once dividends are in arrears for six months, the preference shares have voting rights. This, coupled with the damage to the reputation of the company caused by the late payment, should deter such behaviour on the part of the issuing company.

Immelman (1980) used a narrower definition to avoid this problem. He selected companies that had not paid the preference dividend for two or more consecutive years where the financial position of the company was weak (presumably judged by some subjective means).

3.3.7. COMPANY HAD POOR FINANCIAL PERFORMANCE

Using poor financial performance to select failed companies expands the definition of corporate failure towards the "Economic Failure" and "Financial Failure" end of the spectrum illustrated in Figure 3.1. However, such an assessment is a subjective one. In order to define it more specifically, two routes have been followed in the South African literature.

Firstly, Daya (1977) and Court (1983) combined a subjective assessment of poor financial performance with a subsequent delisting from the Johannesburg Stock Exchange in order to select failed companies. The second approach, applied by, Immelman (1980) and De la Rey (1981), set an objective hurdle of zero or negative profits for two consecutive years as the criterion for failure selection. The latter, however, may not always indicate a failure if, for example, the company is launching a profitable product which is in its pioneer stage.

3.3.8. COMPANY FAILED TO PAY ORDINARY DIVIDEND

Lau (1989) identified five-states of financial distress in her study. On her continuum of financial distress, she considered the omission or reduction of a dividend as the state closest to financial stability. She argued that although a financially stable firm may reduce or fail to declare dividends for non-failure related reasons, empirical studies (Donaldson (1969); Pettit (1972); Dielman and Oppenheimer (1984); and Gentry, Newbold and Whitford (1985)) have shown that a firm that reduces dividends is typically encountering financial distress.

Extending the selection criteria to include situations in which the providers of capital receive no or a lower return expands the definition of corporate failure further towards the "Economic Failure" end of the spectrum. While failure to pay an ordinary dividend may indicate poor financial performance or a cash flow problem, funds may also be withheld in order to invest in projects that will provide significant risk-adjusted returns. It may also be the dividend policy of a company to pay little or no dividends.

For this reason, simply selecting a company as failed based on the fact that it has not declared a dividend in a particular year, as used by De la Rey (1981), may not be correct in all instances. Le Roux (1980) adapted this criterion by selecting companies that had not declared dividends in two successive years. This still suffers from similar problems.

In an additional failure criterion, Le Roux considered dividend policy. A company was considered as having failed if it reduced its ordinary dividend in year two on that of year one and then in year three declared no dividend. A falling dividend payment indicates a change from normal dividend policy. However, this does not consider a possible intentional change in policy or an increase in available profitable projects.

Immelman (1980) addressed the issues of profitability and dividend policy by selecting companies that had ceased to declare a dividend where there had been an indication of a consistent dividend policy prior to the cessation and where the financial position of the company was judged as weak.

3.3.9. TAKEOVER OF A COMPANY COUPLED WITH OTHER CRITERIA

When a company is failing, a healthy company may acquire or merge with the failing company if the healthy company finds synergies between it and the target or thinks it can turn the performance of the failure around (Bulow & Shoven, 1978). The company is considered to have failed because it could not survive alone and could not defend itself against the takeover.

Immelman (1980) considered all such companies as having failed if the takeover occurred one year after the target had published poor financial results. Le Roux (1980) selected a company as a failure if it was taken over after having failed to declare a dividend.

CHAPTER 4
SELECTION OF SAMPLE COMPANIES

Implicit within the definition of corporate failure are a number of symptoms that characterise such a scenario. These symptoms, once identified, can be used as criteria when assessing whether a company has failed according to the applicable definition. These criteria are the starting point for the selection of the failed companies in any empirical corporate failure study. This chapter starts by discussing such criteria, before progressing on to a discussion of the process for the selection of the non-failed portion of the sample.

4.1. DICHOTOMOUS AND MULTIPLE-STATE MODELS

Most models constructed in corporate failure research have used a dichotomous classification technique for classifying firms into one of two states - failed or non-failed. In fact, all studies reviewed in this section, other than the research published by Lau (1987), have used such a two-state model.

Ohlson (1981, 111) argued that investigating a decision problem with only two possible outcomes is not reflective of the rich set of choices confronting investors in the real world. He argued that the lack of consensus on the definition of "failure" in empirical studies has been symptomatic of this.

> "No decision problem I can think of has a payoff space which is partitioned naturally into the binary status bankruptcy versus non-bankruptcy. (Even in the case of a "simple" loan decision, the payoff configuration is much more complex)."

Lau (1987) attempted to address this "continuum of financial distress" by constructing a five-state financial distress prediction model using multinomial logit analysis:

- financial stability,
- omitting or reducing dividend payments,
- technical default and default on loan payments,
- protection under Chapter X or XI of the Bankruptcy Act (in the U.S.),
- bankruptcy and liquidation.

4.2. SELECTION OF FAILED COMPANIES

The sample set of failed companies is selected by applying the failure definition for a particular study to a population of companies. The population from which the sample can be drawn is limited by:

- industry,
- country, and
- time period,

as mentioned in the previous chapter. These factors are discussed below.

4.2.1. SELECTION OF INDUSTRY FROM WHICH TO DRAW SAMPLE

Ohlson (1980) identified utilities, transportation companies, financial services companies (banks, insurance, brokerage, etc.) and industrial companies as structurally different groups. As a result each has a different bankruptcy environment. Platt & Platt (1990) identified such inter-industry variations as a prime reason for poor *ex post* classification results for many failure prediction models.

A possible solution has involved selecting a sample of companies from a single industry. Some studies, that have segregated data in this way, have examined financial institutions (Lane et al, 1986; Looney, Wansley & Lane, 1989), transportation companies (Altman, 1971) and oil and gas companies (Platt et al, 1994) as separate samples.

Within the broad manufacturing/retailing industry in the U.K., Platt et al (1994) found that there were an insufficient number of publicly traded failed companies to segment the analysis. Consequently, many studies have assumed the similarity of companies across manufacturing industries (for example, Beaver, 1966; Blum, 1974; Elam, 1975; Peel et al, 1986; to name a few). This has been despite evidence that the operating and financial characteristics of companies differ between the industries within this broad "manufacturing" environment (Platt, 1989 and Mensah, 1984).

However, Altman et al (1977) found that adding retailers to a sample of manufacturers did not negatively affect prediction results. They proposed that this is because of the similar GAAP used in both of these sectors.

Ultimately, however, the decision on how to segregate a population of companies will be influenced by the availability and accessibility of related information.

Other solutions to the problem of multiple-industry populations are discussed below and in Chapter Five.

4.2.2. PERIOD OF COMPANY SAMPLE SELECTION

Altman (1968, 593) admitted that using a long period over which to sample companies is not optimal. He noted that this is because average ratios tend to shift over time.

> "Ideally we would prefer to examine a list of ratios in time period *t* in order to make predictions about other firms in the following period (*t*+1)."

Altman was not able to achieve this because of data limitations. However, his sample was spread evenly over the twenty year period that he investigated.

Deakin (1972) highlighted an example of a time series distortion in financial ratios through a comparison of his own findings to those of Beaver (1966). Deakin pointed out that the ratio of cash reserves to sales of the failed companies in his U.S. sample taken between 1964 and 1970 were on average significantly higher than Beaver's sample average taken between 1954 and 1964. Deakin went on to suggest that a possible reason for this was the higher interest rates experienced in the U.S. during the late 1960's. Corporations tended to hold more of their cash in reserves to benefit from this interest income. This was in contrast to Beaver's explanation that such high ratios were an indication of cash mismanagement. This is a classic illustration of a problem that exists in selecting a relevant feature subset with which to construct a failure prediction model (discussed further in Chapter Eight).

From a behavioural aspect point of view, users may change the way in which they combine information in their decision making process over a period of time. If these changes are significant, there will be an inconsistency in the information requirements between these periods (Libby, 1975).

Blum (1974) attempted to build a mechanism into his failure prediction model that would allow for it to be continually updated for such changes in trends of information over time. He was not able to verify that the predictive accuracy of this model would improve through the incorporation of such a procedure. Although he could not verify this over the time period that he studied, he noted that over a longer horizon this updating may still be important – it would allow for the inclusion of changes in the macro-economy that might influence the causes of the various business failures.

4.2.3. COUNTRY OF COMPANY SAMPLE SELECTION

Differences in generally accepted accounting practice and economic infrastructures make comparisons between countries difficult.

Platt et al (1994) excluded Canadian firms from their sample of oil and gas companies. Tax regulations and the accounting methods employed in accounting for oil and gas operations differ between companies in the U.S. and Canada. For these reasons, Platt et al decided that the two countries were not comparable within a single failure prediction model.

4.3. SELECTION OF NON-FAILED COMPANIES

In the literature, "non-failed" has been defined, by default, as the converse of the researchers' definition of "failed". In most U.S. and U.K. studies, where bankruptcy has been used for failed company selection, companies that were not bankrupt at the time of the study were used as the population from which to select the non-failed sample (Peel et al (1986), Ohlson (1980), Altman (1968)). Conversely, in other studies where failure has been defined using a broader range of the financial distress spectrum, the definition of non-failed has been relaxed (Blum (1974)). South African studies have taken the "relaxed" stance to this definition.

The term "success" to denote "non-failed" has not been used in the literature, except by Walters (1980, 4). However, Walters even qualified this term by stating that "success is extremely difficult to define without interposing one's subjective evaluation. For this reason, the term "success" has had to be defined in "non-failure" terms".

Two methods have been used for the selection of non-failed companies in corporate failure research. The "paired" approach matches each non-failed company to a failed company based on some predefined criteria. The "non-paired" approach selects non-failed companies independently of the failed company data set.

4.3.1. PAIRED SELECTION

Most early studies selected the non-failed company samples on a paired basis. These studies matched non-failed to failed companies on the basis of one or more of the following criteria:

- Industry and asset size of the company

- Date of failure
- Unobservable factors

(a) Asset size and industry

Beaver (1966, 73) justified the paired approach by arguing that it is necessary in order to provide "control" over factors that might otherwise blur the relationship between ratios and failure.

Beaver controlled for industry sector because, across different industries, the same numerical value of a ratio may imply a different probability of failure.

In addition, empirical evidence has indicated that the variability of the rate of return on assets becomes more stable as company asset size increases (Alexander, 1949). The implication is that large firms have a lower probability of failure, even if the values of their ratios are the same as those of the smaller firm. Beaver (1966, 72) argued that this actually makes larger firms a more critical group on which to perform research

While Altman (1968), Norton & Smith (1979) and Zavgren (1985) controlled for asset size using the value of total assets, Blum (1974) and Elam (1975) used sales and the number of employees to control for this factor. Norton and Smith argued that asset size is preferable to sales as asset size is more stable over time. Blum argued that sales are more reflective of the size of the business, as different businesses need to make different asset investments in order to run the same scale of operations.

(b) Date of failure

Altman et al (1977) and Elam (1975) paired the data of each non-failed company to that of the failed company by collecting data for the same years from both companies.

Platt et al (1994) argued that such a paired sample design controls for the time series distortions of financial data spread over business cycles and varying economic conditions. They also argued that such pairing will increase the degrees of freedom by pooling companies across time.

> "Temporal distortions may arise whenever time series data are analysed as cross-sectional data." (Platt et al, 1994, 493)

Alternative methods to eliminate this temporal distortion are discussed later in this chapter.

(c) Unobservable factors

According to Zavgren (1985, 23), samples can be matched to control for many unobservable factors. This is because there are some factors that are not measurable but should still be controlled for so that the variables are more representative of a homogeneous sample.

4.3.2. NON-PAIRED SELECTION

Non-paired selection has often been justified on the grounds of the minimal impact that it has on biasing the predictive power of the model and the distribution of the final sample selected.

(a) Impact on predictive power of model

Proponents of non-paired sample selection have argued that, while paired design mitigates the disruptive influence of the industry and asset-size factors, it also virtually eliminates any predictive power that these factors may have had (Beaver, 1966, 75). Lennox (1999) noted that a paired approach mitigates the investigation of the affects of industry sector, company size or year of failure on the probability of bankruptcy. In fact, Ohlson (1980) found size to be a significant predictor of corporate failure when included in a forecasting model.

(b) Impact on sample distribution

Lennox found that matching on the basis of company size led to a larger proportion of small companies to large companies in the non-failed sample than the same proportion in the population from which the non-failed companies were being drawn. This is because of the greater probability of failure for a small company.

Similarly, matching samples on the basis of industry type will lead to the inclusion of many companies from "recession-hit" industries.

In addition to the above drawbacks, the methodology used in a particular study may also require the use of a non-paired sampling technique. Deakin (1972) criticised Altman's Z-score model – he noted that it violated the assumption of Multiple Discriminant Analysis (MDA) that the samples be drawn randomly from the population. If paired sampling techniques are used, more complex MDA procedures, which were not followed by Altman, need to be followed.

Zmijewski (1984) demonstrated that the proportion of failed to non-failed companies in a sample should be in proportion to the population survival ratio (failed to non-failed ratio for population over the relevant period). In a paired sample design, where such a ratio is not maintained, certain additional statistical procedures should be applied. The specific procedure will depend on the methodology that used. Platt et al (1994), in their study of the oil and gas industry, used the entire population of non-failed oil and gas companies to make the sample completely representative of the population. This was made possible by the fact that there were only 125 non-failed companies in their population. In this way, there was less concern over Zmijewski's (1984) population ratios problem.

Selection bias may also creep in if only those firms that are clearly non-failed in relation to the sample of failed companies are selected. The non-failed sample should incorporate companies of all levels of financial health. In an attempt to avoid such a selection bias, Atiya (2000) randomly selected a set of non-failed firms across the whole spectrum of financial health - from healthy to borderline.

Inherent to this debate about the merits of the "paired" and "non-paired" approaches, is the consensus that the sample of failed and non-failed companies still needs to be drawn from the same population (Beaver, 1966, 76). For example, each study reviewed in this section drew its sample of failed and non-failed companies from the same time period and same country. Boritz & Kennedy (1995) took this a step further and collected their non-failed company sample so that there was a similar distribution of failed and non-failed companies over the period under investigation.

4.3.3. Basic Considerations

The research is not clear as to which is the optimal method for sampling the non-failed companies. However, there are a few key considerations that need to be taken into account. These include, amongst others, the methodology being employed and the group of companies for which the model, ultimately, will predict failure.

CHAPTER 5
FEATURE SET SELECTION

There are a numerous situations that may result in the demise of a firm. These situations may result in the firm exhibiting various symptoms. In the construction of a corporate failure prediction model, these manifest symptoms need to be identified, measured and analysed.

> "Many unobservable factors influence the vulnerability of an individual firm. These include the unmeasured qualities of the assets, the creative ability of management, random events and the decisions of regulators and courts of law. These factors determine the "tolerance for vulnerability", beyond which the firm will fail. Any econometric model containing only financial statement information will not predict with certainty the failure or non-failure of a firm." (Zavgren, 1985, 23)

Martin (1977) recognised that, while there exist critical factors that cannot be measured, one can find "representative" characteristics of these within the population of companies. The prediction of an outcome for an individual will then be correct only if the representative element in the prediction model dominates the idiosyncratic element.

Measurable predictor (or independent) variables (or features) in failure prediction models fall into one of two categories. These variables can either measure:

- characteristics specific to the firm, or
- the macroeconomic environment in which the firm is operating.

The selection of a feature set to be investigated should not be confused with the process of selecting which of the variables in a set best contribute to the prediction accuracy of a model. Evaluating the information content of different subsets is discussed Chapter Seven.

5.1. CATEGORY 1: FIRM-SPECIFIC PREDICTOR VARIABLES

5.1.1. FIRM-SPECIFIC FINANCIAL DATA

(a) Format of financial data

Blum (1974) compared the predictive ability of financial information in ratio and non-ratio format. He found that ratios were superior predictors, especially one year before failure. In all the studies reviewed in this section, company financial information has been transformed into ratio format before being used in the failure prediction models.

> "These transformations presumably allow more direct comparisons of different size firms and a better picture of a firm's financial position and the interrelationship of the data." (Elam, 1975, 26)

In more recent studies, statistical transformations of data have been executed before including the variable in the prediction model. For example, Ohlson (1980) took the logarithm of the company asset size to scale all sizes between zero and one. In this way, he argued, the effects of outliers on the model were minimised. In addition, the standard deviation and regression slope lines have sometimes been included in the model as predictor variables (Blum, 1974).

(b) Criteria for the selection of financial ratios

Different studies have cited various reasons for selecting a set of features. These reasons have included:

- Popularity in the literature (Beaver, 1966; Altman, 1968; Altman et al, 1977): This reason has most often been cited in the literature. Beaver noted that it may be self-defeating to select such popular ratios because these ratios would most likely have been manipulated by management. Another problem with this approach is that the number of ratios utilised by various studies will increase over time.
- Simplicity of the ratio (Ohlson, 1980).
- Ratios that have performed well in previous studies (Beaver, 1966; Altman, 1968).
- Ratios that are relevant to a particular study (Altman, 1968; Platt et al, 1994).
- Ratios that investors would use and consider as helpful in predicting bankruptcy in reality (Libby, 1975; Norton & Smith, 1979).

- Ratios selected on the basis of their derivation from a theory of business failure, discussed further below (Beaver, 1966; Blum, 1974; Lau, 1987).

(c) Selection of ratios based on theories of corporate failure

> "The products of an accounting system are always surrogates; they are useful only because they represent principals, i.e., the economic events of an entity. This point can never be over emphasised" (Ijiri, 1967, 6)

Blum (1973) took this view. He selected those accounting data variables that he judged to represent the symptoms of the various theories for corporate failure. Although, in using this approach, the complexity of the economic world is reduced to a few accounting scalars, these can be viewed as representing the factors that result in failure in the real world. Blum summarised his factors into: profitability, liquidity and variability. Blum then constructed his Failing-Company Model (FCM) by relating these three common denominators to the cash-flow concept described below.

Beaver (1966) defined the cash-flow concept. In terms of Beaver's framework, the firm is viewed as a reservoir of liquid assets which is supplied by inflows and drained by outflows. The reservoir serves as a buffer against variations in the flows. The solvency of the firm can be defined in terms of the probability that the reservoir will run dry. This concept is discussed further in Chapter Fifteen.

Both Beaver and Blum used this concept for the selection of predictor variables. Table 5.1. below indicates how Blum combined the cash-flow concept with his three failure-related factors in order to select the independent variables for his study.

Concept/Theory	Related Financial Predictor Variable
I. Liquidity:	
A. Short-run Liquidity	
Flow:	1. The "quick" flow ratio[a]
Position:	2. Net quick assets / inventory
B. Long-run Liquidity	
Flow:	3. Cash flow / total liabilities
Position:	4. Net worth at fair market value / total liabilities
	5. Net worth at book value / total liabilities
II. Profitability:	6. Rate of return to common stockholders who invest for a minimum of three years
III. Variability:	7. Standard deviation of net income over period
	8. Trend breaks for net income[b]
	9. Slope for net income[c]
	10-12. Standard deviation, trend breaks, and slope of the ratio number 2 for two years before failure.

[a] cash + notes receivable + marketable securities + (annual sales/12) / (cost of goods sold - depreciation expense + selling and administrative expense + interest)/12
[b] A trend break is defined as any performance by a variable less favourable in one year than in the preceding year, such as a decline from $10,000 to $1,000 from year three to year four before failure.
[c] Slope of "trend" line fitted to the group of observations by the method of least squares.

Table 5.1. The Failing Company Model (Source: Blum, 1974, 16)

As an alternative, Lau (1987) used the "financial flexibility" concept, developed by Donaldson (1969), as her framework for selecting explanatory variables. This concept views the maintenance of a firm's funds-flow balance as the determinant of its solvency. Donaldson identified five "financial flexibility" resources:

- borrowing capacity,
- stock flexibility,
- cost flexibility,
- dividend flexibility, and
- asset disposability.

Lau then selected a financial variable to measure each resource.

(d) Behavioural aspects of using accounting ratios

Libby (1975) recognised that researchers had investigated the predictive ability and behavioural impact of accounting information separately. There had been many studies performed to empirically support the use of certain accounting ratio information in forecasting corporate failure. Hofstedt (1972) investigated the behavioural impact of accounting variations on decision making.

> "Given that the predictive power of the measurements (the accuracy of the signals) and the ability of the decision maker (DM) to use the information (the accuracy of the DM's response to the signal) jointly determine the quality of decisions, it would seem beneficial to use a methodology that examines both factors jointly." (Libby, 1975, 150)

Libby's study was designed to determine whether accounting ratios provided relevant information that could be used by the decision maker within his limitations as an information processor. Using a subset of Deakin's (1972) 14 variable prediction set, commercial loan officers were asked to analyse the ratios and then to predict "failure"' or "non-failure". Individual differences in cognitive judgement are commonly recognised in psychology (Wiggins, 1973). In a situation in which accounting information is optimal for one decision maker but sub-optimal for another, accountants would be forced to decide which user's utility to maximise.

Information was judged by Libby to be optimal if it allowed the users to make the correct predictions. Libby found that the loan officers had a significant predictive accuracy using accounting information. He concluded that the ratio information was utilised correctly by the loan officers.

Libby also investigated whether users may change the way in which they combine information in their decision making process. If these changes are significant, there will be greater consistency in information requirements over consecutive periods than over periods separated by long intervals. He found that the interpretations of accounting data did not vary greatly across time.

5.1.2. FIRM-SPECIFIC NON-FINANCIAL DATA

Non-financial information has been used as predictor data in corporate failure studies (Ohlson, 1980; Peel et al, 1986; Merks, 1986; and Whittred & Zimmer, 1984, amongst others). Such non-financial information has been extracted from a company's annual report. Acquiring such data is often more costly than acquiring financial information.

(a) The lag and changes in the lag of reporting financial statements

> "Infrequency of financial statements is an age-old sign too often ignored. It's axiomatic that the borrower who has had a good year will don his track shoes and speed the statements to your desk to receive the high praise so justly due to him for his managerial acumen. It is just as axiomatic that there is always a reason for not getting the statements on

a timely basis and, when they are received, finding them not comparable to the prior year." (Mitchell, 1976, 5)

Lawrence (1983) found that the issue of the annual reports of failing firms in the U.S. was delayed significantly longer than those of healthy firms. Peel et al (1986) found that this lag was a significant contributor in their logit corporate failure prediction model. In Australia, Whittred and Zimmer (1984) found that there were marked differences between the reporting behaviour of failed and non-failed firms. For example, they found that 75% of non-failed firms released their financial statements within four months, while only 25% of failed companies met this deadline a year before failure. However, they did find that such differences were only significant for firms for a period of up to two years before failure.

U.K. companies, as with South African companies, are subject to penalties for the late issue of financial statements and may have their listing withdrawn as a result. Peel et al (1986) suggested that any delay in the issue of these financial statements would have to be made with good cause.

The literature has suggested a number of possible reasons for this relationship:

- **The Auditing Process**: The auditing process could be particularly problematic and time consuming for firms in "poor" shape (Ohlson, 1980). Whittred and Zimmer (1984) gave examples of these auditor-client delays. These included the query of discretionary accounting changes, management taking exception to auditor's qualifications, as well as expanded audit procedures to cope with the symptoms and risks associated with a failing company.
- **Management Delays**: Management may delay the publishing of financial statements in order to allow time to rectify any perceived deficiencies that will be reflected in the accounts (Peel et al, 1986).

Whittred and Zimmer (1984) measured three different delays in their study. These were the number of days from financial year end to:

- the receipt of preliminary financial statements,
- the date the auditor signs the report and gives his opinion, and
- the date of receipt of the published reports (i.e. printing and mailing).

Other studies incorporating this non-financial variable (Whittred & Zimmer, 1984; Peel et al, 1986) have used the lag between the financial year end and the date the auditor signs the annual report (the second delay described above). This choice was justified by the fact that the impending failure of a company will result in a long delay in the auditing process.

(b) Director resignations and appointments

Director resignations and appointments may occur for a variety of reasons - many not associated with the degree of health of the company. However, Peel et al (1986) suggested that the frequency of resignation and then new appointments may encapsulate the directors' "inside" assessments of the company that are not reflected in the financial information of the company. Resignations may indicate the "abandonment of a sinking ship". Appointments may indicate an attempt or need to strengthen the managerial team.

This variable has been measured by taking the resignations and appointments as a percentage of the total number of directors reported at the financial year end.

(c) Director shareholdings

> "...if directors are viewed as being in a privileged position with regards to price sensitive information, then any changes in their shareholding may signal impending good or bad news which is not necessarily reflected in conventional accounting ratios." (Peel et al, 1986, 7)

Peel et al, also recognised that directors may change their shareholding in their firm's equity for a number of other reasons. In addition, directors are subject to certain statutory and non-statutory controls governing their share transactions, particularly the insider dealing legislation.

This variable has been measured by calculating the change in the ratio of the directors' shareholdings at financial year end to the total issued capital at that date.

(d) Age of the company

Looney et al (1989) found that age was an important predictor in their bank failure prediction model. Beaver (1966) argued that the age of a company will be related to its size, as older firms have had a longer period over which to grow.

5.1.3. FIRM-SPECIFIC CAPITAL MARKET DATA

Beaver (1968) first examined the relationship between a company's share return and the prediction of failure. Blum (1974) was the first to incorporate the return of a company's share price on the stock market as a predictor variable in a multivariate corporate failure prediction study.

> "In an efficient market, stocks already reflect all available information. A forecast about favourable future performance leads instead to favourable current performance, as market participants try to get in on the action before the price jump." (Bodie, Kane & Marcus, 1996, 339)

A problem faced by the firm will typically be reflected in the share price well before it shows up in the balance sheet and income statement (Atiya, 2000, 932). Each period, when the financial report of the company is released, investors will reassess the solvency position of the firm and adjust the market price of the shares accordingly. The direction and magnitude of this adjustment will depend on the size of unexpected changes in the solvency position (Beaver, 1968).

Using this logic as a basis for his study, Beaver tested *ex post* returns of failed firms to see if they were lower than those of non-failed firms (i.e. he tested if failed firms experienced a downward "solvency adjustment" after the release of the annual report). He also attempted to determine the magnitude of the unexpected deterioration in solvency by examining the movement in the share price. Beaver found that companies that failed at a later date did experience lower *ex post* returns, as hypothesised.

If a firm has a higher probability of failure, it will bear a higher risk and investors will demand a higher return. Blum (1974) found that failing companies had higher risk-unadjusted annual returns than non-failed companies.

Atiya (2000) included the stock price volatility, rather than stock price return, in his neural network prediction model. He found that the financial ratio and stock price volatility model significantly outperformed the model using financial ratios alone.

The results of these studies show a degree of correlation between the capital market and the prediction of the failure of a company.

The improvement in the predictive accuracy of models incorporating capital market data has been explained as follows:

- investors use other non-financial data, which are reflected in movements in the share price (Beaver, 1968);
- the stock market reflects factors such as business conditions and insider information that "trickles" through the market (Atiya, 2000).

5.2. CATEGORY 2: MACROECONOMIC PREDICTOR VARIABLES

As stated by Johnson (1970) in his critique of Altman's (1968) seminal study, "ratios to predict failure do not contain [explicit] information about the intervening economic conditions...the riskiness of a given value for a ratio changes with the business cycle."

While empirical research has demonstrated the ability of early warning models to accurately classify failed and non-failed firms in a selected sample (*ex ante*), *ex post* prediction accuracy has been less successful. Platt and Platt (1990) found that six of the thirteen studies that they surveyed had significantly lower *ex post* prediction accuracy than their *ex ante* classification results. They argued that there were two prime reasons for this:

- inter-industry differences between sample companies, and
- temporal variations in economic environments.

The temporal problem arose because few companies failed in a given industry in a particular year. Companies failing in different years confront different pressures from the external macroeconomic environment. Stiglitz (1972) found that management's decision to seek protection under Chapter XI of the U.S. Bankruptcy Code was dependent on management's perception that poor future economic conditions would prevail. The results of Lennox's (1999) study showed that bankruptcy was more likely when the economy moved from boom to recession and that bankruptcy was less likely if the economy was currently in recession. Lennox showed very clearly that an improvement in business confidence led to a decrease in the probability of failure.

Several methodologies have been used in various studies to account for the non-stationarity of variables over the business cycle. These are discussed below.

5.2.1. CONSTRUCTING SEPARATE FAILURE PREDICTION MODELS FOR DIFFERENT PERIODS OF THE BUSINESS CYCLE

Mensah (1984) examined a model that was built for distinct periods of the business cycle. He found that different sets of predicting variables were significant at different stages of the business cycle, providing evidence of non-stationarity.

Lau (1987) controlled for this variation by constructing a model for predicting failure in a specific year, namely 1977. Data availability severely limits this approach. In addition, a model constructed for a particular year may not have any significant application for other years.

Looney et al (1989) concluded from their evaluation of the misclassified banks in their study that misclassifications were largely caused by changes in the external economy. They concluded that models for failure prediction need to be rebuilt regularly in order to incorporate these changes.

5.2.2. PRICE LEVEL-DEFLATED FINANCIAL INFORMATION

A number of studies have examined the effects of economic fluctuations on prediction models by comparing models built on historic cost financial data and models built using financial data deflated by the general price level in accordance with the applicable accounting standards (Platt et al, 1994 and Norton & Smith, 1979). Deflators convert nominal values into real values by removing the results of external forces, such as inflation.

Empirical evidence of the effect of inflation on the accuracy of failure prediction models has been inconclusive. Norton & Smith (1979) found that bankruptcy models based on historic cost and general price level data did not perform significantly differently. Platt et al (1994) noted that serious queries have been raised with regard to the methodology used by Norton and Smith. In contrast, Short (1978) found that the prediction accuracy including general price level data was significantly better.

Mensah (1984) found that the prediction accuracy of models using general price level data greatly improved the classification accuracy of failed companies (Type I error). No improvement in the accuracy of the classification of non-failed companies was found (Type II error).

Platt et al (1994) attempted to refine the method for deflating financial data. They removed temporal bias through the deflation of each individual financial data type with a variable-specific economic measure. For example, debt was deflated using interest rates and assets were deflated using the movement in the oil price (as it is the oil price that determines the value of the assets held by the oil companies that formed the sample in their study).

Using this refined method, Platt et al noted that by removing the impacts of the external environment, the varying pressures that a company may be facing were eliminated. Deflators, thus, removed valuable information from the failure prediction process.

5.2.3. Inclusion of Economic Predictor Variables in Model

Lennox (1999) captured relative changes in business confidence using a variable constructed from the Confederation of British Industry (CBI) Quarterly Industrial Trends Survey. In this survey, the CBI publishes the results of a questionnaire asking respondents about their optimism in relation to the U.K. economy.

Platt et al (1994) included macroeconomic variables, which were specific to the oil and gas industry under examination, in their prediction model. These included interest rates and oil prices. Wood and Piesse (1987) reported that interest rates are an important determinant of exogenous abnormal returns.

Platt et al, found that adding macroeconomic variables into a model that already included nominal financial ratios did not add any predictive ability to the model. They argued that the nominal financial data incorporated both their own dynamics and the effects of external economic factors, as movements in the macro-environment will affect the nominal financial performance of a company. In other words, there existed a degree of multicollinearity between the external variables and the nominal ratios in their study.

In contrast, economic variables made significant independent contributions to the prediction of failure in models based upon deflated (real) financial ratios. Platt et al argued that financial information sets with temporal bias that has been removed, do not contain information pertaining to the economic effects of the period under investigation. Hence, the macro-variables add explanatory power to the model. They argued conceptually, and showed empirically, that there are no significant predictive advantages to a model that uses real financial ratios with external variables over one that is based on non-deflated (nominal) financial ratios alone.

5.2.4. INDUSTRY-SPECIFIC PREDICTOR VARIABLES

Lau (1987) used the industry debt-equity ratio to adjust each company's individual debt-equity ratio. She achieved this by dividing the company ratio through by the industry ratio.

By comparing financial ratios to industry averages, Platt and Platt (1990) attempted to alleviated the inter-industry bias. However, Platt et al (1994) found such industry variables to be unnecessary because the scope of their study only included a single industry (oil and gas).

Other studies, such as the one performed by Lennox (1999), have included an industry dummy variable to represent each industry within the sample. Lennox found that the dummy variable was important because different variables impacted each industry in a different manner.

CHAPTER 6
DATA PREPARATION

Once the samples of failed and non-failed companies have been selected, and the overall set of features has been chosen, the data for all the features for each sampled company need to be collected and processed. This chapter starts with a discussion of some of the key considerations in this collection process. Thereafter, different methods for data pre-processing are addressed.

6.1. DATA COLLECTION

The time period, the country in which the study is being performed and the type of information required, all impact on the time and resources spent on data collection. In addition, consideration still needs to be made for:

- the number of years of data that needs to be collected for each sampled company;
- from where the required data can be sourced; and
- how to deal with financial reporting lags and variations in accounting policy.

6.1.1. YEARS FOR WHICH DATA IS COLLECTED

(a) Failed companies

Corporate failure studies start by collecting data for all the sampled failed companies for each year prior to the companies' defined dates of failure.

The number of years of data collected for each company is usually dependent on how far in advance the model under construction is attempting to predict the impending failure. These forecast periods have ranged between eight years (Blum, 1974), five years (Beaver, 1966; Altman, 1968; Deakin, 1972; Elam, 1975; Altman et al, 1977; Sharma & Mahajan, 1980; Court, 1991); and two years (Truter, 1996) in the published literature. Data availability may limit this forecast period (Elam, 1975).

(b) Non-failed companies

The period for which data are collected for the non-failed companies in an empirical study, has depended on whether a paired or non-paired sample technique has been employed by the researcher.

In the **paired technique**, data has been collected for the non-failed company for the same years as its failed company pairing (Blum, 1974; Altman et al, 1977). In some studies, however, companies were not paired on the basis of date of failure. In these studies, the researchers collected data for the non-failed companies randomly so that the failed and non-failed companies were distributed similarly over the period under study (Beaver, 1966; Altman, 1968).

The approach for data collection when a **non-paired sample technique** has been used, has varied.

- Commonly, the non-failed company sample has been constructed in such a way as to obtain a sample ratio of failed to non-failed years of data that mirrors that of the population (Platt et al, 1994). Certain methodologies require such an approach.
- Ohlson (1980) noted that for large sample sizes in the manufacturing/ retailing sector, the cost of collecting a number of years of data for each non-failed company can be extremely costly. Instead, Ohlson collected a single year of data for every non-failed company randomly from within the period under investigation.
- Frydman, Altman & Kao (1985) and Lau (1987) selected a randomly available number of non-failed companies for their samples. In contrast to methodologies that assume a population representative ratio, the methodology employed by Frydman et al, recursive partitioning analysis, does not require any specific population ratio to be maintained within the sample.

6.1.2. COLLECTION OF ANNUAL REPORT DATA

The majority of corporate failure studies have used some form of financial data in the construction of their failure prediction models. Such financial information has seldom been collected directly from the annual report, as was done by Ohlson (1980). In most cases, the author has extracted the data required from a database that summarises the required financial information.

In older studies, periodicals served as the means for storing such data. For example, in the U.S., the Moody's Industrial Manual was used by many studies, including Beaver (1966), Altman (1968), Blum (1974) and Elam (1975), as the source for their data. Peel et al (1989) used the Stock Exchange Official Handbook to obtain similar data in the U.K. More recently, computer databases have been established to store such information. Lau (1987) and Zavgren (1985), among others, used the COMPUSTAT tapes to extract the data required for the construction of their models in the U.S.

In some studies, the authors have applied a new methodology to an existing data set. Deakin (1972) and Whittred & Zimmer (1984) obtained their data sets from Beaver (1966) and Altman & Izan (1982), respectively.

Where more specific information, such as accounting policies and certain non-financial information (for example, reporting lag and director shareholdings) are required, the actual published reports may need to be obtained (Ohlson, 1980; Platt et al, 1994). The cost of such a process needs to be weighed against the benefit that such additional data will contribute to the prediction accuracy of the model.

6.1.3. PITFALLS IN THE DATA COLLECTION PROCESS

(a) Reporting lags

> "The time lag between a company's accounting financial year end and the date the annual accounts are actually published (and changes in this lag from the previous year) might vary *inter alia* in sympathy with any changes in the "news content" of the accounts." (Peel et al, 1986, 6)

Ohlson's (1980) study was the first to explicitly consider the date at which a set of financial statements is issued. The annual financial report of a company is not issued at the financial year end. In fact, as discussed Chapter Five, the lag between the financial year end and the date of issue of the annual report has been shown to become longer in companies with a higher risk of failure. Such reporting lags would result in:

- the availability of financial information for forecasting failure being made available later in the case of failing companies, and
- the possibility of information being disclosed months after a firm petitioned for bankruptcy or some other definition of failure was already met. The result would be to overstate the classification accuracy of the model by using information only available after-the-fact.

Lawrence (1983) found that a significant number of failed firms in his sample incurred delays in releasing their annual reports for the final year before failure. Approximately 22% of his sample of 58 industrial firms in the U.S. had filed for bankruptcy before they released their annual reports. Ohlson found the comparative rate on his sample to be 17%. Ohlson also noted that all studies prior to his own ignored this reporting lag problem. Hence, the accuracy of models prior to Ohlson's study may suffer from overstated classification accuracy.

Some of the studies that have followed Ohlson's study have explicitly considered this problem. For example, Lennox (1999) checked to ensure that none of his sample had issued financial statements after bankruptcy.

(b) Variations in accounting method employed

Different industries may employ different accounting methods. Even within an industry, there may be various policies from which a company may choose. For example, in the oil and gas industry in the U.S., companies may choose between the full cost and successful efforts accounting methods. Many researchers have noted the impact that these variations can make on model construction.

Platt et al (1994), in their study of the oil and gas industry, controlled for this problem by introducing a dummy variable to distinguish between the successful efforts and full cost methods.

Altman et al (1977) spent significant time adjusting their financial data set so that it was consistent in incorporating several of the most recent accounting modifications of that time:

- all non-cancellable operating and finance leases were capitalised;
- reserves of a contingency nature were included in equity and income was adjusted for the net change in the reserve for the year;
- minority interests and liabilities were netted against other assets to allow for a truer comparison of earnings with assets generating the earnings;
- goodwill and intangibles were deducted from assets and equity because of the difficulty in assigning an economic value to them; and
- capitalised research and development costs, capitalised interest and certain deferred charges were expensed to improve comparability and to give a better picture of actual fund flows.

Atiya (2001) made the point that various databases summarise their financial data using different accounting methods or by grouping items, such as intangibles, in a different manner. This would make the extraction of information from multiple sources problematic.

(c) Data availability

Sharma & Mahajan (1980) recognised that the variables they were able to use were limited by data availability. Methods for dealing with missing data are dealt with in the following sub-section.

6.2. DATA PRE-PROCESSING

The process of pre-processing data, in its most basic form, entails converting the data collected into a ratio format. The merits of the use of ratios have been discussed in Chapter Five. Most importantly, ratios allow for the rescaling of different nominal financial values into a range that does not arbitrarily assign a single value significantly more weight than the next. In other words, turnover should not bear more significance in a model than the cash balance simply because it is numerically larger.

Data pre-processing refers to analysing and transforming the predictor variables in order to minimise noise, highlight relationships, detect trends and adjust the distribution of the variable so as to optimise the methodology being employed (Kaastra & Boyd, 1996, 220).

This can involve a number of processes. The manner in which raw data is analysed and transformed during each process is largely dependent on the requirements of the researcher and, more importantly, the requirements of the methodology being employed.

6.2.1. DATA FILTERING

It is important to assess the distribution of the data. The removal of certain observations may be important in creating a more uniform and representative data distribution. The detection of outliers is an example of this.

6.2.2. DATA TRANSFORMATION

This is the process of converting the data into a form that has more information content for failure prediction purposes.

Altman et al (1977, 32) expressed certain variables in a logarithmic form in order to reduce the outlier effects and in order to adhere to certain statistical assumptions.

Studies that have employed neural network methodologies have, historically, addressed the issue of data transformation in more detail than studies that have employed other methodologies. This is because machine learning techniques rely on pattern-recognition. Through data transformation, the patterns between the "input/target" pairs can be made more distinct (this is sometimes referred to as "feature extraction"). Therefore, researchers employing machine learning techniques will focus a greater degree of attention on this step.

Zirilli (1997, 43) defined feature extraction as "the process whereby the raw input data is transformed into input/target pairs". In machine learning terminology, the predictor variable data set relating to a specific company can be viewed as a vector of "inputs". The classification of the company related to this "input" vector is the "target" variable (i.e. failure or non-failure).

In machine learning, these "input/target pairs" are an important concept. The input/target sets bear within them the patterns that represent the distinction between failed and non-failed. It is therefore imperative that such data sets are complete in respect to information content. However, simultaneously, they should include a broad enough range of information so that the model can predict classification in situations that are similar, but not identical, as those used to create the model.

Typically, a form of normalisation has been used in order to transform the data (Etheridge & Sriram, 1997; Welch et al, 1998; O'Leary, 1998). Normalisation is dealt with in more detail in Chapter Fifteen.

6.2.3. FEATURE SUBSET SELECTION

The most common way of pre-processing the data set into a form that better explains the distinction between failed and non-failed, is through the selection of a relevant subset of the features. This is discussed in more detail in the following chapter.

CHAPTER 7
FEATURE SUBSET SELECTION

The inclusion of a full set of features has, in almost all cases, a detrimental effect on predictor model accuracy. This is because of redundant features (which are dependent on other features) and irrelevant features (which add noise to the information input into the model). Furthermore, the more features there are, the more likely that some feature will randomly fit the data, hence, increasing the probability of over-fitting. Removing these features leads to both improved accuracy and a clearer description of the relevant factors that contribute to the failure process.

In addition, by identifying fewer features, the cost of collecting data for the implementation of the final model is reduced.

In the classic corporate failure studies reviewed in this report, a number of standard statistical approaches for predictor variable subset selection have been used. These have been summarised by Eisenbeis, Gilbert & Avery (1973) in their article entitled "Investigating the relative importance of individual variables and variable subsets in discriminant analysis".

This chapter starts with a brief description of these procedures, Principle Component Analysis (PCA) and evolutionary search techniques. The chapter then progresses on to an analysis of the types of variable subsets that have been selected in the international published literature. Finally, the chapter concludes with a brief discussion of how a machine learning approach impacts on the feature subset selection process.

7.1. THE RELATIVE SIGNIFICANCE OF INDIVIDUAL VARIABLES

7.1.1. METHODS FOR INDEPENDENT VARIABLE SELECTION

> "The procedure of reducing a variable set to an acceptable number is closely related to an attempt to determine the relative importance within a given variable set." (Altman et al, 1977, 36)

With this in mind, many of the early variable selection procedures attempted to calculate the importance of each variable individually.

There are a number of methods that have been used for evaluating individual predictor variables (examples of studies that have employed these methods have been included):

- **Forward stepwise selection methods** (Altman et al, 1977; Norton & Smith, 1979): This method begins by selecting the variable which, individually, best explains the distinction between the failed and non-failed classes in the sample. The explanatory power of each variable can be evaluated using a number of different measures. For example, in the case of Norton & Smith, the multivariate F-ratio was used. The selected variable is then paired with each of the remaining variables in order to determine which pair provides the best explanatory power. The procedure continues until all variables that meet a minimum criterion for inclusion have been included.

- **Backward stepwise methods** (Altman et al, 1977): This method works conversely to the forward stepwise method. Variables are excluded, one at a time, from a starting point that includes the full variable set. The variable that reduces the explanatory power of the variable set the least when excluded from this complete set, is excluded. Once again, the F-ratio is commonly used to evaluate the explanatory power of the variable set.

- **Scaled vector test** (Deakin, 1972; Blum, 1974; Altman et al, 1977): A scaled vector measures the relative contribution of each variable to the model. It is calculated by multiplying each discriminant coefficient by the square root of the appropriate variable in the variance-covariance matrix. Larger values for the scaled coefficient indicate greater variable explanatory power.

- **Separation of means test** (Altman et al, 1977): This measures the relative contributions of each independent variable to separating the means of the failed and non-failed groups of companies. This method was suggested by Mosteller & Wallace (1963) and supported by Joy & Tollefson (1975).

- **Univariate F-statistic** (Altman et al, 1977; Norton & Smith, 1979): The univariate F-statistic measures the ability of each individual ratio to predict failure.

7.1.2. PROBLEMS WITH INDIVIDUAL VARIABLE SELECTION METHODS

Norton & Smith (1979) described the results of the **univariate F-statistic, forward and backward stepwise procedures** as "optimal" but not "maximal". They argued that a maximal solution would require testing every possible subset of ratios that could be used. In only examining the contribution of a single variable at a time, a ratio that has insignificant explanatory power alone, but that may contain important information when combined with other ratios, will be excluded from the model. However, an exhaustive search would require testing N^2-1 possible combinations (where N is the number of variables in the full feature set). This may make computational time prohibitively expensive and in certain cases impossible (Altman et al, 1981).

Blum (1974) made the point that the relative importance of each variable cannot be measured reliably with a **scaled vector**. This is because the model coefficients are highly unstable when the variables included are highly correlated with each other (multicollinearity). In fact, in the case of perfectly correlated variables, the model coefficients are assigned arbitrarily. Thus, the relative weights of these variables are not representative of their relative importance.

Norton & Smith (1979) and Altman et al (1977) noted that there may be a high degree of instability in the ratios selected using these methods. Norton & Smith admitted that such instability may be caused by multicollinearity, a problem not addressed by these methods.

7.2. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) or factor analysis is a method that identifies which ratios constitute the principal independent dimensions of the data (Zavgren, 1985, 23). Each independent dimension is orthogonal (bears no collinearity) to the other dimensions. In this way, the problem of multicollinearity can be avoided.

Libby (1975) was the first to use factor analysis to identify independent sources of variation. In this way he reduced the number of variables and highlighted the prominent financial characteristics of the data set. He identified the five following independent sources of variation:

- profitability,
- activity,
- liquidity,
- asset balance and
- cash position.

Zavgren (1985) also used factor analysis to identify the principal independent dimensions of financial statement data. Zavgren noted that PCA will consistently identify the same orthogonal dimensions each time it is run on a data set.

In South Africa, Le Roux (1980) and Court (1993) have used factor analysis for predictor variable selection in corporate failure studies.

7.3. EVOLUTIONARY SEARCH TECHNIQUES

Genetic Algorithms (GAs) are stochastic search techniques that can search large and complicated spaces by applying ideas from genetics and evolution (Davis, 1991; Holland, 1975; Goldberg, 1989). Since there are many hundreds of features that can be generated from a company financial report alone, a technique that can search for an optimal solution in a very large search space is well suited to feature subset selection in corporate failure prediction.

A number of studies have employed GAs with great success in selecting a feature subsets (Kingdom & Feldman, 1995 and Piramuthu, 1999). In particular, these studies have found that GAs are successful when used in combination with neural network classifiers.

An adaptation of this technique, Population-Based Incremental Learning (PBIL), is implemented in this report. There does not appear to be prior research in which PBIL has been used in a corporate failure study.

7.4. EXAMPLES OF FEATURE SUBSETS SELECTED IN THE LITERATURE

Some studies have not used a variable subset selection method. For example, Ohlson (1980) merely chose variables that had been popular in studies preceding his own. Other studies, however, have used empirical feature subset selection techniques.

7.4.1. ALTMAN'S Z-SCORE FEATURES

Within the literature, the features used by Altman (1968) in his seminal work are considered to be classic with respect to this field of study.

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \qquad (7.1.)$$

For low values of Z in the above Z-score model there is a greater probability of bankruptcy. The variables selected for Altman's Z-score discriminant function have been justified as follows:

- **X1 - Working Capital / Total Assets**:
  This ratio measures the net liquid assets relative to the total capitalisation of the firm. According to Altman, if a firm is experiencing consistent operating losses, it will have shrinking current assets in relation to its total assets.

- **X2 - Retained Earnings / Total Assets**:
  The age of the company is implicitly considered in this ratio. Altman argued that a younger firm will show a low retained earnings figure because it has not had time to accumulate profits. Thus, with a lower ratio, younger firms will have an increased chance of being classified as failures – this is the case in reality.

- **X3 - Earnings before Interest and Taxes / Total Assets**:
  Altman described this variable as the true measure of the productivity of a firm's assets, ignoring any tax or leverage factors.

- **X4 - Market Value of Equity / Book Value of Total Debt**:
  This measure shows how much the market value of equity can decline before the liabilities exceed the assets and the firm becomes insolvent.

- **X5 - Sales / Total Assets**:
  This ratio illustrates the sales generating ability of a firm's assets. Altman claimsed that it is a measure of management's capability in dealing with competitive conditions.

7.4.2. FEATURE SUBSETS SELECTED IN OTHER STUDIES

Appendix B.2. summarises the ratios selected as significant predictors in 29 of the most widely cited international corporate failure prediction models reviewed in this study. These models also cover a broad range of different classifier methodologies employed in this field of research (see Appendix B.1.). It is clear that a broad range of variables have been used across different studies. The South African studies reviewed in this report have a similar trend in the range and type of ratios selected for model construction.

While some studies have gone no further than simply selecting a set of features for input into a prediction model, other studies have taken the additional step of justifying their variable choice. Extensive discussions in this regard can be found in the work done by Libby (1975), Altman et al (1977) and Lennox (1999). Such analysis has sought to extract the reasons and extent to which certain variables and business factors contribute to the failure process.

For example, Lennox found that, "to summarise, a company is most likely to go bankrupt when it is unprofitable, highly leveraged and has cash flow problems". Although such results may seem similar to those of other studies, Lennox also found that the effects of cash flow and leverage are non-linear. For this reason, Lennox concluded that a non-linear model should outperform a linear model – a finding that further motivates for the use of non-linear induction learning techniques.

7.5. FEATURE SUBSET SELECTION IN A MACHINE LEARNING APPROACH

7.5.1. THE IMPORTANCE OF FEATURE SUBSET SELECTION IN THE IMPLEMENTATION OF AN INDUCTIVE CLASSIFIER

As noted in the introduction to this report, machine learning can only be a successful approach to a problem where there is actually knowledge contained within a sample. Then, using the correct induction technique, an algorithm can distil this knowledge from the sample data by attempting to discover the relationships between features and classes. Thus, the success of the induction process is heavily dependent on the selection of the optimal subset of features.

There are a number of additional considerations that need to be taken into account when applying inductive learning classifiers (when compared to the classical statistical approaches). These mainly relate to the concept of generalisation (discussed further in Chapter Nine). In addition, another practical consideration is the greater computational time required for the iterative process of inductive learning.

Both generalisation and computational resource requirements are impacted directly by feature subset selection.

7.5.2. FURTHER DISCUSSION OF FEATURE SUBSET SELECTION FOR INDUCTIVE LEARNING

This chapter simply serves as a summary of feature subset selection techniques as they have been applied to corporate failure research in the reviewed literature. Feature subset selection as it relates to inductive learning is reviewed in detail in Chapter Sixteen.

CHAPTER 8
CLASSICAL STATISTICAL CLASSIFICATION TECHNIQUES

Corporate failure research has been a very popular field of study amongst academics and practitioners for the past four decades. As no consistently accurate and reliable method for corporate failure prediction has yet been found, and as the economic and social implications of failure remain a problem, new methodologies are continually being tested in this field. Many of these techniques have been drawn from other fields of research that have problem domains thought to be comparable to that of corporate failure prediction.

Each technique, both new and old, brings with it its own advantages, as well as assumptions and pitfalls.

The basic assumption of these models is that firms can be classified into distinct groups – typically a dichotomous grouping of failed and non-failed, although studies have been performed using more classes (Lau, 1987) (see Appendix B.1.). Accordingly, firms are characterised by a classification variable, y, such that:

- $y_i = 0$ if the i-th firm is a failure
- $y_i = 1$ if the i-th firm is a non-failure.

Classical statistical methods, such as discriminant analysis and logit/ probit analysis, were initially used for the dichotomous classification procedure. Increasing progress in the field of business failure led to the introduction of techniques such as Recursive Partitioning Analysis (RPA), the Cox Proportional Hazards model, as well as a number of neural network and other machine learning techniques. The various methods that have been used in some of the key studies in this field are summarised in Appendix B.1. The inputs into each of these classifiers are the features selected as described in the previous chapter.

This chapter contains brief explanations of the most commonly applied classical statistical techniques. The basic mechanics of each technique are described before looking at the assumptions, limitations and advantages of the specific technique. This discussion is not intended to be exhaustive. The following chapter reviews the application of neural networks and other machine learning methodologies to this field of research.

8.1. UNIVARIATE STATISTICAL METHODS

Beaver (1966) was among the first to forecast corporate failure using an empirical method and his study is considered a milestone in this area. He used a univariate statistical approach. This approach evaluated each ratio in terms of how it, alone, could be used to predict failure. In an attempt to minimise misclassifications, Beaver calculated a cut-off score for each ratio.

8.2. MULTIPLE DISCRIMINANT ANALYSIS (MDA)

The univariate method was later criticised by Altman (1968) and Libby (1975) who recognised that business failure can be caused and affected simultaneously by many different factors. Based on this criticism, the application of a multivariate statistical method was a logical progression for the field.

Multiple Discriminant Analysis (MDA) is a multivariate analytical technique first employed by Altman (1968), and later employed and "tweaked" in a large number of other studies.

8.2.1. DESCRIPTION OF THE MECHANICS OF MDA

Consider that any firm, *i*, is characterised by a vector, X $(x_{i1}, x_{i2}, x_{i3}, \ldots x_{in})$, which contains all the features that relate to firm *i*. These features form the predictors in the failure prediction model. MDA can be used to estimate a discriminant function that combines the predictor variable vector with a coefficient vector, A, such that the variance between the failed and non-failed firms is maximised.

MDA can combine the predictor variables in a linear or quadratic manner. The linear combination of the variables with coefficient vector, A $(a_1, a_2, a_3, \ldots a_n)$, provides a Z-score for each firm described by (Dimitras et al, 1996):

$$Z_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_n x_{in} \tag{8.1.}$$

Where:

- Z_i is the Z-score for firm *i*, and
- $\{x_{i1}, x_{i2}, \ldots, x_{in}\}$ are the *n* independent variables of firm *i*.

A cut-off score is then calculated. A firm scoring below this level is expected to fail. Altman (1968) simply observed the scores of the failed and non-failed firms in his sample in order to ascertain the critical score value that minimised his classification error. Deakin (1972) improved the accuracy of Altman's subjective critical cut-off point by using a probabilistic classification rule. He also considered the costs of misclassification.

The probabilistic rule can be illustrated with the following figure. The shaded area represents the overlap in classes. The intersection of the distributions would constitute the classification cut-off level.

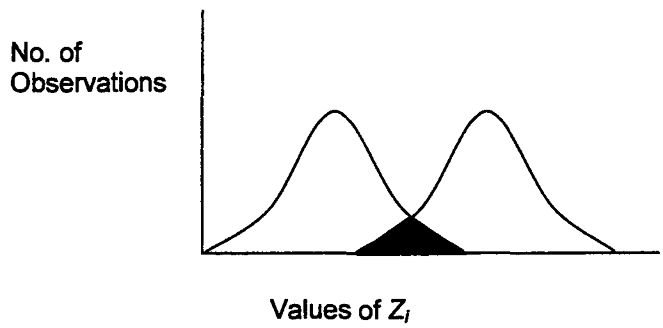


Figure 8.1. Distribution of Discriminant Values for Two Groups (Source: Elam, 1975, Exhibit 2)

8.2.2. Assumptions of MDA

The application of MDA to corporate failure prediction suffers from a number of limitations. In fact, one finds that MDA models are often criticised for ignoring these limitations. For example, the seminal work of Altman (1968) drew much criticism from the academic community, with scathing articles published by Johnson (1970) and Joy & Tollefson (1975).

Each new study employing MDA has sought to improve on a criticised area of a preceding study. Nevertheless, none of these attempts seem to have accomplished higher statistical results than those attained in Altman's initial work. Moreover, in the majority of cases, the practical application of these models has tended to be difficult due to model complexity (Neophytou & Molinera, 2001, 4).

The criticisms of these models have generally revolved around the assumptions of MDA and how these have been violated or simply ignored. The assumptions of MDA can be summarised as follows:

- Explanatory variables are assumed to be distributed within each group (failure and non-failure) according to a **multivariate normal distribution**. Nevertheless, the variables typically used in bankruptcy studies, especially financial ratios, are not normally distributed (Eisenbeis, 1977).

- Linear MDA assumes that each group (failed/non-failed) has **a different mean but an equal dispersion matrix**. If this assumption does not hold, a quadratic structure should be used (Eisenbeis & Avery, 1972).

  Many studies have ignored this assumption, for example Altman (1968) and Deakin (1972). In their ZETA analysis, Altman et al (1977) tested whether a linear or quadratic structure was appropriate (H_1 test, derived from Box (1949)). They found that, even where dispersion matrices were not equal, linear MDA outperformed quadratic MDA on the holdout samples. They concluded that linear MDA is robust in relation to the violation of this underlying assumption.

  However, Eisenbeis and Avery indicated that the use of linear MDA, when the dispersion matrices are unequal, tends to cause more classifications into the group with the larger dispersion. Furthermore, they indicated that "for given differences in group means, the predictive power of the linear rules relative to quadratic rules decrease as the difference between the group dispersions increase" (1972, 38). Norton & Smith (1979) also found this to be the case.

- The sample of failed and non-failed companies is assumed to be **drawn at random** from their respective populations. However, the matched-pairs technique violates this assumption (Altman et al, 1981).

These are some of the reasons why MDA models have been criticised. However, MDA still forms the basis of comparison for new methods that are applied to this field of study (Frydman et al, 1985; Lennox, 1999).

8.2.3. FURTHER LIMITATIONS OF MDA

Deakin (1972, 176) pointed out that in applying the model *ex ante*, one cannot be sure of how many years into the future the company will fail. In addition, one cannot obtain a probability estimate for failure. Finally, the constraints of discriminant analysis prohibit the derivation of discriminant functions where a firm could belong to more than one group (failed or non-failed) over time (e.g. company failing and then rectifies to become non-failure).

8.3. LOGIT AND PROBIT ANALYSES

Ohlson was the first to apply logistic regression or logit analysis in a failure study. He defined his research problem as:

> "Given that a firm belongs to some prespecified population, what is the probability that the firm fails within some prespecified time period?" (1980, 112)

8.3.1. DESCRIPTION OF THE MECHANICS OF LOGIT/PROBIT

Conditional probability models derive the probability of a dichotomous (or polytomous) classification by calculating coefficients for each predictor variable. These coefficients can be interpreted as the effect of a unit change in the predictor variable on the probability of the dependent variable classification. A cumulative probability distribution function is necessary in order to constrain the predicted values within the [0,1] boundaries of a probability distribution.

In these models the probability, P, of a firm i failing, given a firm's vector of predictor variables, X_i, is:

$$P_i = f(a_0 + A.X_i) \quad (8.2.)$$

Where $f\,(a_0 + AX_i)$ is a cumulative probability distribution function giving the probability of the failure of firm i given its predictor variable matrix, X_i, and the coefficient matrix of the model, A.

Logit and probit models differ only in the cumulative distribution functions used. Probit uses a standard normal distribution function where:

$$f(a_0 + A.X_i) = \int_{-\infty}^{a_0 + AX_i} \frac{1}{(2\Pi)^{1/2}} e^{-z^2/2} dz \quad (8.2.)$$

while logit uses a cumulative logistic function:

$$f(a_0 + AX_i) = \frac{1}{1 + e^{-(a_0 + AX_i)}} \quad (8.3.)$$

The difference between the two functions is showed diagrammatically in the figure below. The logit distribution differs from the normal distribution used in the probit model only in being slightly more platycurdic. Zavgren (1985) found no significant difference in the estimation results for the logit and probit models.

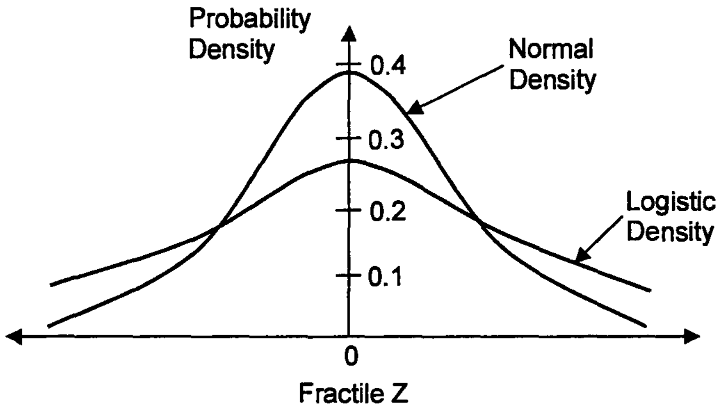


Figure 8.2. Standard Normal and Logistic Density Functions (Source: Altman et al, 1981, Figure 1.6.)

8.3.2. Comparative Advantage of Probability Models over MDA

Logit and probit analyses are not models that are designed to find an "optimal" frontier by trading off one type of error against another. They simply work out a probability schedule of failure and then classify a company accordingly. This is in contrast to MDA which seeks to satisfy optimality conditions, such as minimising misclassification errors (Ohlson, 1980, 126).

The empirical calculation of a probability of failure is a major advantage in the application of logit and probit analyses. For example, as Chesser (1975, 38) noted, the non-compliance of a borrower does not mean that he will completely default on his loan, but rather that some "work-out" agreement will have to be arranged. An assessment of the probability of such an event makes differential adjustments, such as risk-premiums on interest rates and loan indentures, possible.

It is possible to generate probabilities of failure when using discriminant analysis. This involves a subjective assessment of the probability associated with a particular discriminant score. However, when the sample of the population under study contains a strongly non-representative proportion of the group of failed or non-failed companies in relation to the population, the subjective probability criterion of MDA will be biased.

Since most discriminant analysis studies use equal-sized matched samples, most studies will suffer from this bias. Martin (1977, 258) provided evidence that the probabilities obtained from a discriminant function may be inaccurate, despite high classification accuracy.

Likewise, the conditional probability model requires the likelihood function to be weighted so that the sample proportion of bankrupt companies is approximately equal to that of the population - otherwise all coefficient estimates are biased (Lennox, 1999).

8.3.3. Results Achieved by Logit/Probit

Most previous studies have argued that, in practice, the explanatory powers of probit and logit models are similar to that of MDA (Zavgren, 1985; Peel et al, 1986; Platt et al, 1994). However, Lennox (1999) performed tests for omitted variable bias (La Grange Multiplier tests) and heteroskedacity. In so doing, and by allowing for the non-linearity of variables, the logit and probit models significantly outperformed the MDA model.

Nevertheless, the research has not been conclusive on the superiority of performance of any of these models.

8.4. RECURSIVE PARTITIONING ANALYSIS (RPA)

A commonality between univariate models, MDA and logit and probit analysis is the parametric quality of the linkage between the explanatory variables and the groupings. A host of statistical errors related to the parametric assumptions of such models have been cited as rendering their results somewhat problematic.

> "RPA is a computerised, nonparametric technique based on pattern recognition which has attributes of both the classical univariate classification approach and multivariate procedures." (Frydman et al, 1985, 269)

8.4.1. Description of the Mechanics of RPA

This method builds a binary classification tree with a specific rule associated to each node on the tree (a node is a point where the tree branches). These are usually univariate rules - a certain financial characteristic is given a cut-off point that minimises the cost of misclassification of a firm into the incorrect group.

In order to classify a company, one descends the tree, based on the rules at each node, until finally identifying the group membership of the specific firm. This method also calculates the associated probability of group membership.

The figure below illustrates one of the classification trees constructed by Frydman et al (1985). The letters "NB" represent "not bankrupt" and "B" stands for "bankrupt".

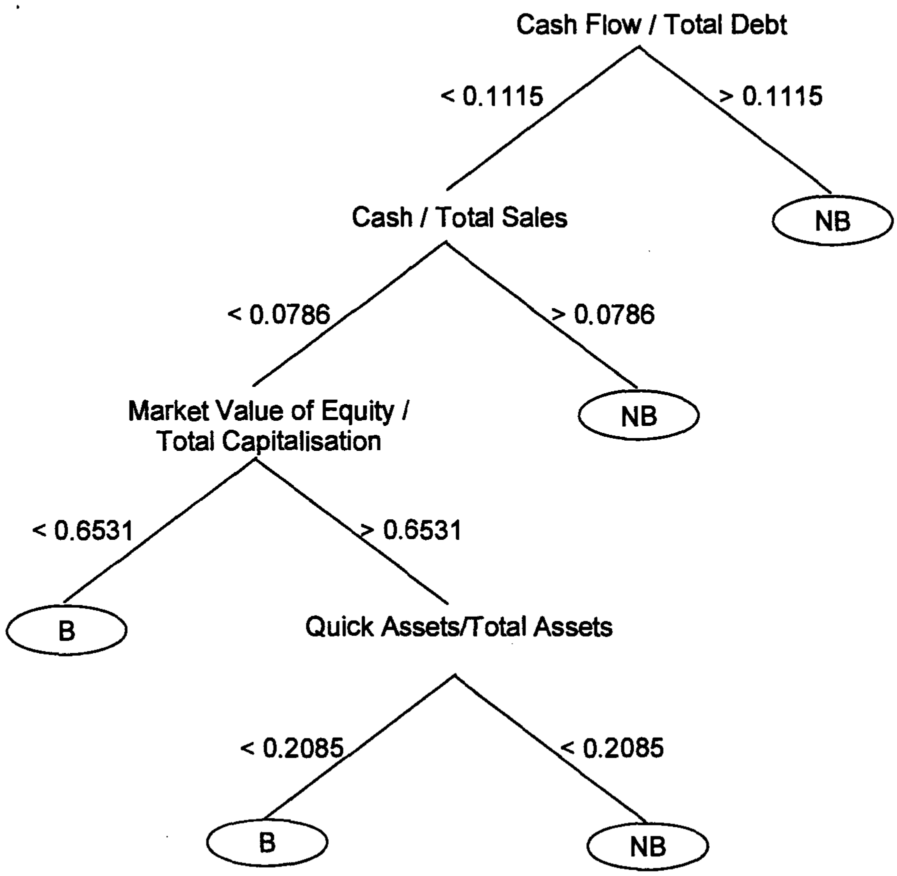


Figure 8.3. The RPA1 tree (Source: Frydman et al, 1985, Figure B.1.a.)

The inputs required to build an RPA failure prediction model are:

- financial information for each company,
- actual group classifications,
- prior probabilities of failure and
- the costs of misclassifying firms into the incorrect group.

8.4.2. Comparative Advantage of RPA over Other Models

Frydman et al (1985) first employed RPA in corporate failure prediction. They also compared the RPA model to a model derived using MDA and found no significant difference in prediction accuracy. However, RPA did have a number of other advantages over such comparative models.

RPA is a nonparametric method with the assumptions that groups are discrete, non-overlapping and identifiable (as with MDA). The fact that there are no distributional assumptions with RPA is an attractive feature of this technique. For example, the lower bound of zero on ratios like sales over assets automatically skews any possible normal distribution for such a ratio. The normal distribution of this input variable is an assumption underlying most parametric techniques.

Another advantage of RPA is that the binary classification tree helps to explain the causes of failure for a particular firm. In models where a linear combination of variables is used, it is not easy to determine which specific factors have resulted in the failure of a firm. However, in RPA, the extent of the contribution of individual variables is not completely unambiguous (Frydman et al, 1985).

8.5. COX PROPORTIONAL HAZARDS MODEL

This model was developed by Cox (1972) and applied extensively in the biomedical field, especially with heart transplant and cancer data (e.g. Crowley & Hu (1977)). Lane et al (1986) first applied it to bank failure prediction.

8.5.1. Description of the Mechanics of the Cox Proportional Hazards Model

The model uses a hazard function, $h(t)$. This function calculates the probability of failure in the next instant, given that the company was alive at time t. Using this probability function, the technique can be used to calculate how much time there is until the probable failure of a firm.

8.5.2. Comparative Advantage of the Cox Model over Other Models

The Cox model has few underlying assumptions. In addition, it has the capability to specifically incorporate the time until failure into the modelling process (Lane et al, 1986).

It is important to consider the time dependence of the variables in designing a sample to be used in the Cox model. The Cox model makes a prediction of time to failure assuming information is available on a continuous basis. However, companies do not publish financial information on a continuous basis - only annually or at quarterly or semi-annual reporting intervals. The assumption of this model is that the values of the financial ratios for a particular company remain constant between these reporting intervals.

CHAPTER 9
NEURAL NETWORKS AND MACHINE LEARNING CLASSIFIERS

The previous chapter reviewed a number of classical statistical approaches to corporate failure prediction. However, over the last decade or so, extensive research has also been performed in this field using neural network applications.

The origins of neural networks as a field of study lie in brain theory. However, nowadays such methods are simply seen as another way of using a function to model data through an algorithm - "learning" the patterns within that set of data. This is the also the case with modern machine learning techniques. In fact, *IEEE Transactions on Neural Networks* makes no distinction between classical neural network papers and those using approaches such as kernel methods.

This chapter begins by describing the development and basic structure of neural networks from the perspective of brain theory. A discussion linking neural networks to statistical data modelling and inductive learning methods follows. This discussion incorporates a motivation for the use of the inductive kernel methods used in this study. Finally, the chapter concludes with a brief review of the different neural network and machine learning techniques applied to corporate failure prediction in the literature.

Firstly, however, the chapter begins with a discussion of a fundamental consideration of any classification model – generalisation.

9.1. GENERALISATION

Henery (1994, 6) noted that classification by learning can have two distinct meanings:

- **Unsupervised learning**: In the one situation the classification task may require establishing the existence of previously undetermined classes within the data based on a set of observations. This method can be used as a technique to explore the underlying structure of the data (Boritz & Kennedy, 1995, 504).
- **Supervised learning**: Alternatively, the classes may be known. The aim is then to establish a rule whereby an unknown observation can be allocated to one of these

classes. This is the case in this study and, as such, all references to learning in this report are making reference to supervised learning.

Inductive supervised learning can be viewed, simplistically, as a process of mapping a set of inputs to a known set of classes.

Figure 9.1. Basic illustration of supervised inductive learning process

The parameters of the universal function in Figure 9.1. are "trained" so that inputs are mapped to outputs. However, the desired model is not one that maps this relationship exactly. Rather, the objective of the process is to produce a function that can generalise well when a set of previously unseen inputs are presented to the algorithm.

Generalisation is defined by Vonk et al (1997, 8) as the "property ... whereby a [model] is able to provide a correct matching of output data to a set of previously unseen input data." The concept of generalisation is a critical one and is illustrated in the diagram below.

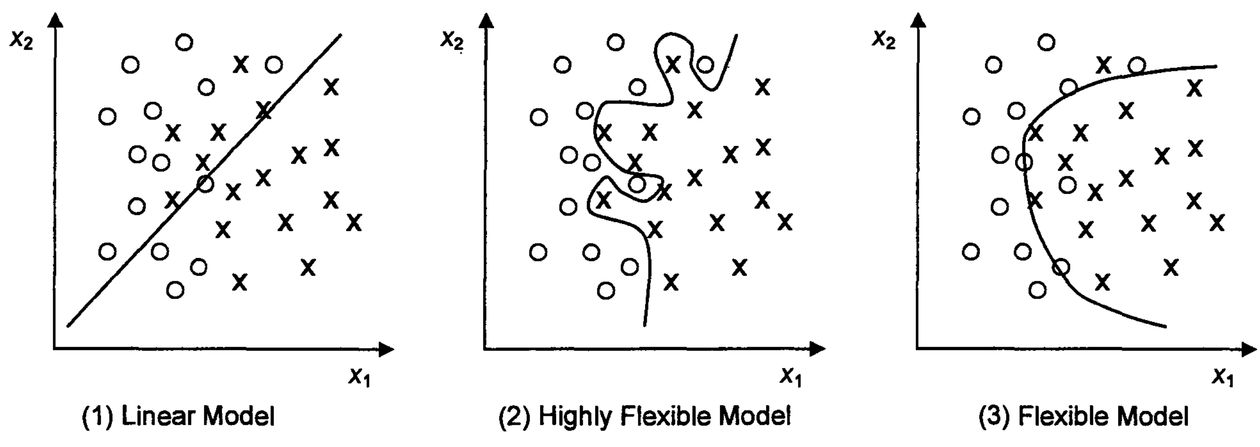


Figure 9.2. These graphs show a dichotomous sample (shown by crosses and circles) with a two feature vector (x_1 and x_2). Each solution boundary varies in complexity. (Source: Bishop, 1995, 11)

In Figure 9.2. above, model (1) is a simple linear boundary and gives poor separation.

Model (2) achieves perfect separation of the training data but with a boundary too complex to generalise a "best estimate" for an "unseen" data point (over-fitting). In fact, this model estimates the "accidental" properties of the data set, including noise and error. Both models (1) and (2) have poor generalisation.

Model (3) gives a good separation of training data with a boundary that can better generalise on "unseen" data points. It predicts the simpler underlying trends which have better predictive power for unseen data.

The concepts of generalisation and over-fitting are referred to throughout this report.

9.2. NEURAL NETWORKS FROM THE PERSPECTIVE OF BRAIN THEORY

Neural networks are able to map complex non-linear patterns between sets of input data and the given output for that data.

> "In practice, neural networks are ... useful for classification and function approximation/mapping problems which are tolerant of some imprecision, which have lots of training data available, but to which hard and fast rules (such as those that might be used in an expert system) cannot easily be applied." (Sarle, 2000)

9.2.1. HISTORY

Historically, many concepts in neural computing have been inspired by the study of biological networks. The "fathers" of neural networks based their theories on the original theories of how the brain was thought to function. Haykin (1994, 2) described a neural network as a "massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use". He identified two respects in which artificial neural networks resemble brain theory:

- Knowledge is acquired by the network through learning.
- Interneuron connection weights, known as synaptic weights by neuroanatomists, are used to store the knowledge.

The historical developments in neural computing can be summarised as follows (Arbib, 1995 and Olmsted, 1998):

1873 **First biological neural networks**:

The first biological neural network was presented by Alexander Bain (1818 – 1903) of the United Kingdom. His work was inspired by the neuroanatomists who were able, for the first time, to see the true extent of the interlocking fibers of the brain. In his book "*Mind and Body. The theories of their relation*" (1873), Bain first described memory as a set of nerve currents weaker than that produced by the original stimulus.

1906 **Further anatomical studies**:

In Spain, Santiago Ramon y Cajal completed detailed anatomical studies of the brain revealing the particular structure as a network of neurons. In England, Charles Sherrington published physiological studies on synapses, the junction point between nerves.

1943 **Modern neural network theory**:

McCulloch and Pitts developed the first modern neural network theory in their book "*A Logical Calculus of the Ideas Immanent in Nervous Activity*". They united the studies of neurophysiology and mathematical logic. Their model of the neuron as a "threshold logic unit" was built on the anatomy of the neuron by Ramon y Cajal and the excitatory and inhibitory synapses of Sherrington.

1949 **Computer simulation**:

Donald Hebb published "*The Organisation of Behaviour*", work relating neural networks to computer simulation.

1958 **Perceptron**:

Frank Rosenblatt published work developing the theory of the perceptron.

1986 **Backpropogation**:

Rumelhart, Hinton and Williams describe the popular backpropagation algorithm, the first in a series of neural networks to be applied widely across various disciplines.

However, the more recent advancements in the area of pattern recognition have been from the perspective of neural networks as extensions of conventional statistical techniques. These are discussed in the next subsection of this chapter.

9.2.2. EXPLANATION OF FUNCTIONING

When viewing neural networks from the perspective of brain theory, they can be seen to be composed of several layers of computing elements called nodes that are connected together to form a network.

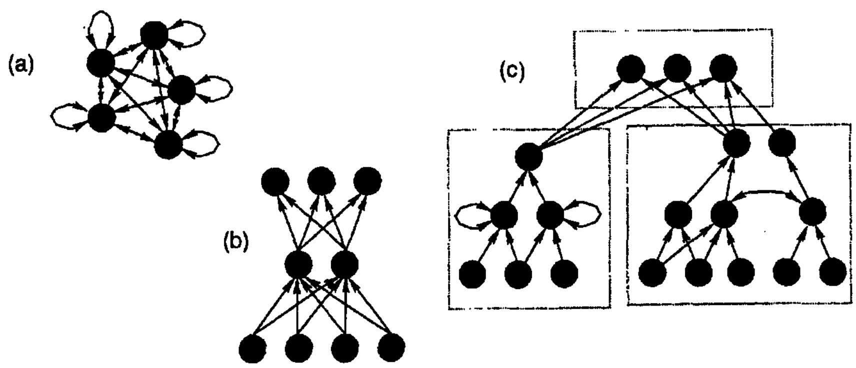


Figure 9.3. Various types of architectures (Plunkett et al, 1997, 2)

(a) A fully recurrent network
(b) A three-layer feedforward network
(c) A complex network, consisting of several modules

Each node is a simple processing unit (likened to a neuron) that receives external inputs or an input signal from other nodes. After processing the signals through a transfer function, it either outputs a final result or a transformed signal to other nodes. This process can be likened to the excitatory or inhibitory inputs a neuron receives. The strength of the influence that one node exerts on other nodes is determined by the connection or "synaptic" weight.

Different neural networks can be characterised by their network architectures (number and connection of nodes), transfer functions, or by any of the other parameters that need to be estimated before the network can be used for prediction purposes.

Diagrammatically, the functioning of an individual node can be expressed as in Figure 9.4. This type of node is known as a perceptron and forms the base unit of a multiple layer perceptron (MLP) network, historically a popular form of neural net.

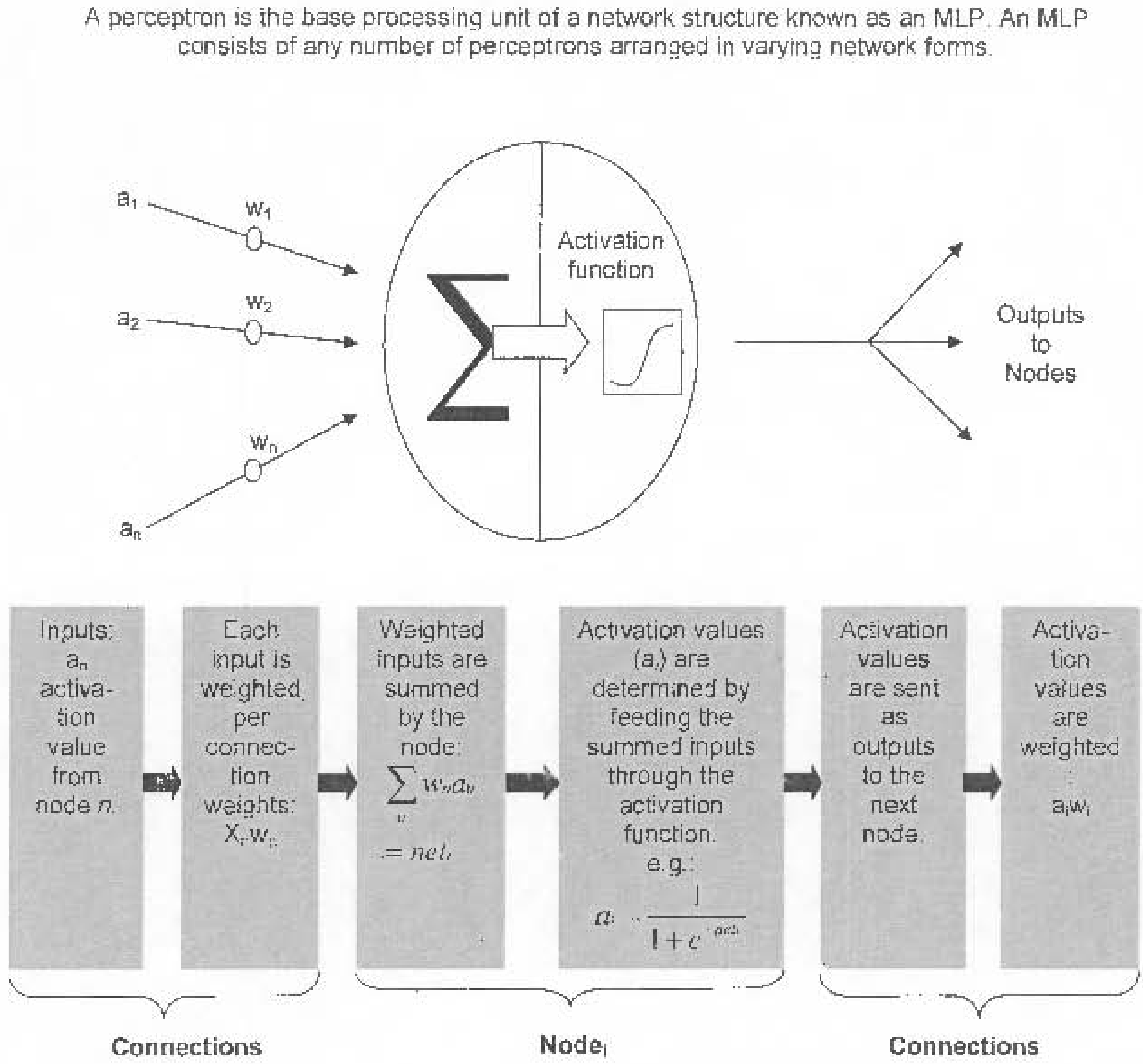


Figure 9.4. The functioning of a single perceptron. A neural network is made up of numerous nodes in varying layers.

The weights between nodes determine the degree to which one node exerts influence over the next. The process of determining these weights (w_i in Figure 9.4.) is called training. The training phase is a critical step in the application of neural networks and is discussed below in the comparison between neural networks and other machine learning techniques.

9.3. NEURAL NETWORKS AND MACHINE LEARNING FROM THE PERSPECTIVE OF CLASSICAL STATISTICAL DATA MODELLING

9.3.1. INTRODUCTION

> "There are several phases that an emerging field goes through before it reaches maturity, and computational finance is no exception."

This is the opening line in the introductory article by Adu-Mostafa and Atiya (2001, 653) to the publication of *IEEE Transactions on Neural Networks*, entitled "Special Issue on Neural Networks in Financial Engineering". The authors then go on to debate the sustainability of neural networks in computational finance based on the quality, novelty and relevance of their results in this field.

Many authors have set about critically assessing the introduction of neural networks into business applications. In an article entitled "Neural Networks: Forecasting breakthrough or passing fad?" Chatfield questioned the "hype" surrounding the claims that neural networks had heralded a "new era in the evolution of forecasting and decision support systems" (1993, 1). He noted that in order to justify this hype, a proportionate amount of effort needed to be expended on neural network research as had historically been expended on traditional statistical techniques.

In the decade since Chatfield published his sceptical views, much research has been performed using neural networks. In fact, the more recent advances in the area of pattern recognition have been from the perspective of neural networks as extensions of conventional statistical techniques.

> "It is a sign of the increasing maturity of the field that methods which were once justified by vague appeals to their neuron-like qualities can now be given a solid statistical foundation." (Bishop, 1995)

The following discussion explains both neural networks and machine learning based on these foundations (Greene, 2004).

9.3.2. STATISTICAL DATA MODELLING

Machine learning algorithms are data driven. Typically, a function is constructed that can be used to "explain" the information content of a sample. If the function is sufficiently smooth (or "regular") it should have the ability to generalise and can be used for forecasting using previously unseen data. This has been discussed above.

The roots of machine learning, thus, lie in classical statistical data modelling, particularly in the work of Fisher (1952). Such early work applied parametric models to the data (for example, normal distributions) and "learning" was limited to estimating the parameters of such distributions.

Gradually statistical modelling began to take account of the fact that reliable prior knowledge of model distributions may not be available. Hence, non-parametric approaches were born. These methods use highly general "universal" models and tune these models to fit the data. However, with an increase in the degrees of freedom of the model, there is an increase in the risk of "over-fitting".

(a) Neural networks

Neural networks are able to create a "universal" function by using a smooth nonlinear function (such as the sigmoid function illustrated in Figure 9.4.) in the hidden nodal units, and adjusting the parameters using a process such as "backpropagation".

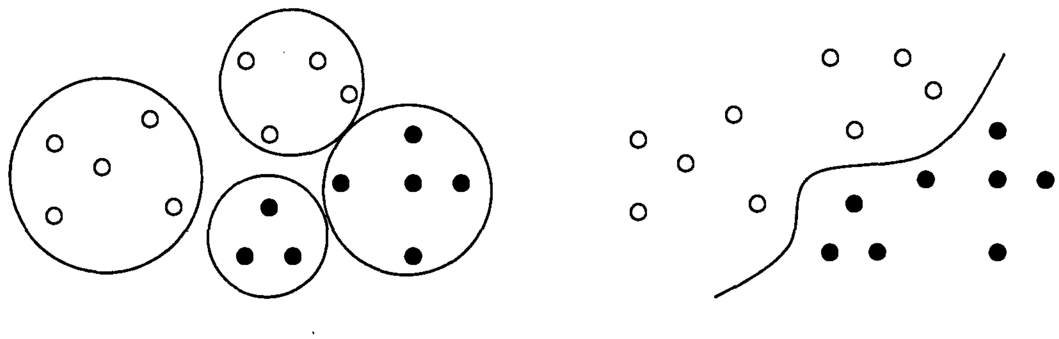
However, this is a difficult nonlinear optimisation process with multiple sub-minima for the loss function and no guarantee of finding the optimal global minimum. With backpropagation, or any other gradient-based optimiser, one has to be content with finding a (hopefully) "relatively good" sub-optimum point.

Despite all the research that has applied such networks, there remains no clear guideline on what the optimal structure of the network should be (such as how many layers, how many hidden units, etc.). Therefore, the experimenter needs to perform a great deal of tedious trial and error experimentation. Nevertheless, with good intuition, or a bit of luck, remarkably good results can be achieved.

In addition, if, when the training process begins, all the weights are set to zero (see Figure 19.4.), the algorithm may remain trapped at its starting point and never progress further. Therefore, in the interests of symmetry breaking, these weights need to be set to small random values prior to commencing the training process. This randomness is likely to lead the algorithm to a different local optimum each time that it is run.

(b) Kernel methods, including SVM and KRR

In an effort to eliminate some of the mystery and simplify the optimisation process, local Radial Basis Functions were substituted for the sigmoidal hidden units. These basis functions could be located *a priori* on the basis of the distribution of the data. An example, where a Radial Basis Function (RBF) has been placed at the centre of the clusters of data, has been illustrated below. The type of result that would be achieved with a Multi-Layer Perceptron (MLP) neural network has been included for comparative purposes.



(A) RBFs **(B) MLPs**

Figure 9.5. The difference between RBFs and MLPs

An extreme version of this is to place a radial basis function on every data point. This leads to the "probabilistic neural network" and to kernel methods in general. The problem of over-fitting with such an approach can be resolved by regularisation. This is discussed in detail in Chapter Twenty.

The advantages of kernel methods are that, unlike the hit-and-miss approach of neural networks:

- this method is virtually self-designing (a kemel on each data point, versus the choice of hidden weights in the case of neural networks), and
- training has a unique minimum (kernel methods only train the alpha values, or "output weights", while the multiple parameters of a neural network that need to be trained result in numerous sub-minima).

Support Vector Machines (SVMs) and Kernel Ridge Regression (KRR), both kernel methods, are very similar and almost identical in performance. The SVM uses a well-defined quadratic optimisation process, while the KRR alpha weights are determined by the least squares solution of an over-determined system of linear simultaneous equations. The difference between them is in the regularisation process. This has been discussed in detail in Chapter Twenty.

9.3.3. TERMINOLOGY

Trigueiros & Taffler have provided a glossary that explains certain neural network terminology through equivalent traditional statistical terms. While this may be simplistic in certain areas, it is useful for the purpose of demystifying neural networks and relating them to more commonly used techniques.

Neural Network Terminology	Statistical Modeling Terminology
Neural Network	Model
Synapses, weights, connectivities	Coefficients of the model
Inputs	Independent variables
Outputs	Dependent variables
Outcome or target	Expected value
Node	Logistic regression
Hidden layer	Intermediate set of logistic regressions
Learning	Coefficient estimation
Supervised learning	Regression, discriminant analysis, etc.
Unsupervised learning	Principal components analysis, etc.
Architecture	Model description
Convergence	In-sample performance
Generalisation	Out-of-sample performance

Table 9.1. Glossary comparing neural network and statistical terminology (Source: Trigueiros & Taffler, 1996, 355)

9.4. THE USE OF MACHINE LEARNING AND NEURAL NETWORKS IN CORPORATE FAILURE PREDICTION

Research studies in which machine learning has been applied to bankruptcy prediction started in 1990 and continues today. The vast majority of such studies employ neural networks.

9.4.1. APPLICABILITY OF MACHINE LEARNING AND NEURAL NETWORK METHODS

There are a number of reasons why a non-linear approach should be superior to a linear approach:

- Atiya (2000) argued that there are "**saturation**" effects in the relationships between the financial ratios and the prediction of default. For example, if the earnings/total assets changes by an amount of 0.2, from -0.1 to 0.1, it would have a far larger effect on the prediction of failure than it would if that ratio changed from 1.0 to 1.2.

- Atiya (2000) argued further that there are "**multiplicative**" factors that are non-linear as well. For example, the potential for failure of a firm with negative cash flows is amplified if it has larger liabilities. The reason is that a highly leveraged firm will experience an increased difficulty in borrowing money to finance its deficits. This has been discussed further in Chapter Five.

Wong, Bodnovich & Selvi (1997) found that the integration of neural networks with other technologies, such as expert systems, robotics, or decision support systems, improved their applicability in addressing various types of problems. The use of neural networks in practical business applications has grown in popularity since the early 1990's. *Business Week* (1992) and *The Economist* (1995) described such successful implementations in a variety of financial and business contexts - including market analysis, bond rating and credit evaluation (Piramuthu, Raghavan & Shaw, 1998). Currently, several of the major commercial loan default prediction products are based on neural networks. For example, Moody's *Public Firm Risk Model* (www.moodysqra.com) is based on a neural network methodology.

The review of the literature dealing with corporate failure prediction using machine learning techniques follows.

9.4.2. Brief Review of Literature

(a) Seminal work

The first attempt to use neural networks to predict corporate failure was made by Odom & Sharda (1990). In their study, a three-layer feedforward network was used and the results were compared to those of a multi-variate discriminant analysis model. They tested the use of different proportions of failed to non-failed firms in their training sample on their model's predictive ability. Neural networks were found to be more robust in both training and test results.

Many different neural network studies followed this pioneering research – many using the same data set as Odom & Sharda in an attempt to find the best network structure and optimisation algorithm (Coleman, Graettinger & Lawrence, 1991; Rahimian, Singh, Thammachote & Virmani, 1993).

Tam & Kiang's research (1992) has had a great impact on the use of neural networks in their application to corporate failure prediction. Based on Tam (1991), they provided a detailed analysis of the potential advantages and limitations of such classifiers. They compared neural networks to statistical methods such as linear discriminant analysis, logistic regression, k-nearest neighbours and decision trees, another machine learning technique. They found that neural networks were generally more robust and accurate.

(b) Comparative results

Much research has been performed using different neural networks over the last decade. Some studies have found that they outperform many of the traditional techniques discussed in the previous chapter (Fletcher & Goss, 1993; Brockett, Cooper, Golden & Pikatong, 1994; Wilson & Sharda, 1994; Sharda & Wilson, 1996). Other studies have found no such improvement (Altman et al, 1994; Podig, 1995; Kerling, 1996).

Altman, Marco & Varetto (1994) criticised the "black box" approach of neural networks. As highlighted earlier in this chapter, there are many parameters that go into the construction of a neural network classifier, as well as intuition and luck on the part of the researcher. Many results published over the years should be viewed with scepticism. Leshno & Spector (1996) evaluated a wide range of different neural networks and concluded as follows:

- The prediction accuracy of the model depends on the sample size used for training.
- Different learning techniques have significant effects on both model fitting and test performance.
- Over-fitting problems are associated with a large number of iterations used in training.

Yet, many studies have used small samples, (for example, Fletcher & Goss (1993) use only 18 in their sample) and have applied different techniques and iterations without sound theoretical justification.

As discussed above, the state-of-the-art machine learning techniques employed in this study are not as volatile and risky as the traditional neural network approaches.

(c) Optimisation algorithms employed

Most studies use the backpropagation algorithm (Odom & Sharda, 1990; Salchenberger, Cinar & Lash, 1992; Tam & Kiang, 1992; Wilson & Sharda, 1994). As discussed earlier in this chapter, this algorithm has a number of critical limitations. Piramuthu et al (1994) found that different optimisation algorithms do have affects on the performance of neural networks.

Kernel Ridge Regression, employed in this study, is trained with a least squares solution to an over-determined system of linear simultaneous equations. The advantages of such an approach over backpropagation have also been addressed above.

CHAPTER 10
EVALUATING FAILURE PREDICTION MODELS

A classification model is of little practical value if its expected accuracy is not measured and, in fact, compared to the accuracy achievable with other competing models. There are two approaches to measuring such accuracy. Firstly, the classifier's error rate can be estimated. This is the approach adopted in this study. Such an approach is simple to measure and interpret. Another approach entails measuring the significance of the model itself using various statistical techniques.

This chapter reviews both approaches but focuses on the "estimation of error rates" approach because of its application in this study.

It is important to keep the primary goal for the construction of a failure prediction model in mind: that is to find the feature subset and induction algorithm that performs best on data previously unseen by the model.

10.1. ESTIMATION OF ERROR RATES

In testing the accuracy of a classifier, it is commonly accepted that error rates tend to be biased if they are estimated from the same set of data as what was used to construct the model. At the extreme, if a model is sufficiently complex, it will be possible for it to map all the given data and obtain 100% classification accuracy. However, such complex models will perform poorly when applied to unseen data. This is the phenomenon of over-fitting and is discussed in detail in the previous chapter.

10.1.1. VALIDATION TECHNIQUES

While the term "over-fitting" is associated with the field of machine learning and neural networks, this bias created by the intensive searching of a solution space is a problem inherent in any empirical research. Eisenbeis (1977, 893) cited this as one of seven key pitfalls in the application of discriminant analysis. He noted that using all the data to evaluate the accuracy of the MDA model constructed with such data, "leads to a biased and overly optimistic prediction of how well the rules would perform in the population".

Even Altman (1968, 600), in his seminal research, noted the importance of secondary sample testing and its appropriateness in bankruptcy forecasting validation.

This has been applied in a number of different ways in the literature.

(a) Hold-out samples

A common solution to this problem is to split the sample of data into two random and independent training and testing (or holdout/validation) data sets. Various models are then constructed (or "trained" in the case of inductive learning) by minimising the appropriate error function defined with respect to the training data set. The performances of the different models are then compared by evaluating the error function using the independent validation set (Bishop, 1995, 372). This is known as the "hold out" method.

Lachenbruch and Mickey (1968) showed that such an evaluation method performed better when assessing the accuracy of a model on unseen data than training and testing the model with the same data. The predicted and true classifications on this holdout or test data give an unbiased estimate of the error rate of the classifier. A t-test can then be used to test the significance of the results of the holdout sample against the original model constructed with the training data set.

This approach has been used in both classical statistical and machine learning approaches.

Altman (1968) used five replication techniques to select his holdout sample:

- random sampling,
- choosing every other firm starting with firm number one,
- choosing every other firm starting with firm number two,
- choosing firms in the first half of his sample
- choosing firms in the second half of his sample.

Elam (1975) recognised that there is a loss of efficiency in splitting that sample in this way. As the full sample is not used to construct the optimal classification model, there is a trade-off in the selection of the holdout sample size. If the holdout sample selected is too large, then their may be insufficient data with which to train the model. Conversely, if the holdout sample is too small, there may be insufficient holdout sample data to effectively test the model constructed. However, with large data sets this is not a major problem.

(b) Cross-validation

In practice, as is the case in this study, the availability of cases with which to train and test the models may be severely limited. Under such a scenario, the researcher may not be afforded the luxury of keeping aside part of the data set for validation.

Cross-validation (Stone, 1974, 1978; Wahba and Wold, 1975) has been used extensively in the literature to circumvent such a limitation. It has also been used extensively in the application of machine learning algorithms in the field of corporate failure prediction (Zhang, Hu & Patuwo (1999) made reference to its numerous applications in this field of study).

This method is a form of "resampling" (Ahn, Cho & Kim, 2000). The data set is split into *X* mutually exclusive portions. Training is conducted on *X*-1 of the portions and tested on the remaining portion. This training is repeated using all possible combinations of *X*-1 portions, each time testing the model on the omitted portion. The error of the model is determined by averaging the error rates across all tested segments.

A more specific example, where *X*=5, is described below.

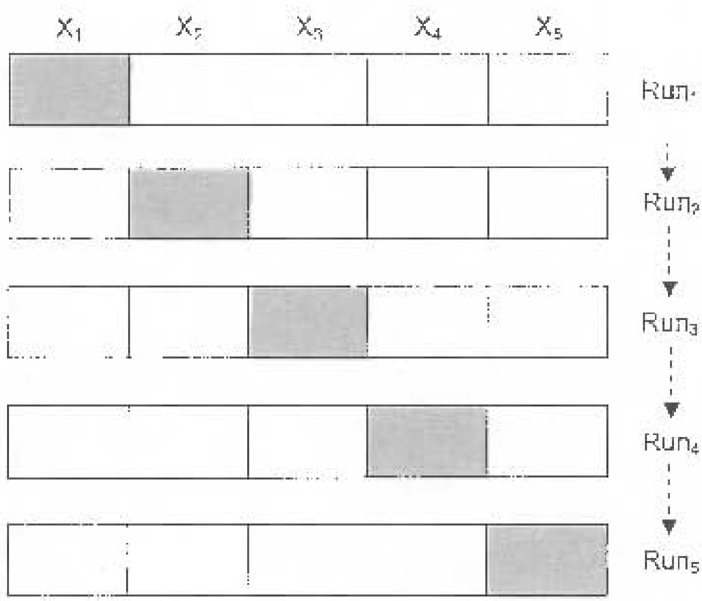


Figure 10.1. A schematic illustration of the partitioning of data into five segments for five-fold cross-validation.

The above model is trained five times, each time using a version of the data set in which one of the segments (shown shaded) is omitted. Each trained model is then tested on the data from the segment which was omitted during training, and the results averaged over all five models (Bishop, 1995, 375).

In this way the use of available data is maximised (as all cases are used in both testing and training the model), while still maintaining independent testing and training data splits.

There are, however, two major drawbacks to this method:

- Bishop (1995, 374) made the point that as the training process for each model has to be repeated *X* times (if X-fold cross-validation is used), there may be circumstances in which this could become computationally intensive and time consuming.
- Greene (2001) noted that this method is only valid if each of the independently trained classifiers is sufficiently similar to warrant the averaging of their performances. For example, the use of a multi-layer perceptron, which depends on random effects in its training, can be too unstable to justify such averaging.

In corporate failure prediction, even where the data is split into two independent subsets, the accuracy of the model for periods after that included in the study cannot be known with any degree of certainty. This is because of changes in the measurement, use and trend of company financial information over time. This has been discussed in detail in Chapter Five.

Peel et al (1986) took a separate hold-out sample of companies with the most recent set of published accounts. In this way, they attempted to test the inter-temporal predictive power of their model.

(c) Leave-One-Out (LOO) validation

Lachenbruch (1967) suggested an almost unbiased validation technique that maximises the available sample. He suggested the construction of a number of models, holding out one isolated observation at a time. This isolated observation is then tested for accuracy. The individual observations' classification accuracy is then cumulative over the entire set of models constructed. Many studies employ this method (among others, Elam, 1975; Norton & Smith, 1979; Zavgren, 1985; Lennox, 1999).

Using the example illustrated in Figure 10.1. above. LOO can be viewed as cross-validation that has been taken to the extreme. The learning system, consisting of N cases, is trained N-1 times, each time omitting only a single case. The overall error rate is then, simply, the sum of the errors made.

Not only does this maximise the data with which the model is trained, but it also allows for the rapid update of the model with each new sample case added. However, this is computationally intensive.

10.1.2. Types of Prediction Errors

A type I error is the case in which a prediction model incorrectly classifies a failing firm as healthy. A type II error, in contrast, is the situation in which a healthy firm is incorrectly classified as failing. These errors can be illustrated by the following classification matrix:

		Actual Outcome	
		Failure	Non-Failure
Predicted Outcome	Failure	Correct Prediction	Type II Error
	Non-Failure	Type I Error	Correct Prediction

Table 10.1. Classification Matrix

10.1.3. Costs of Error Types

Overall error rates ignore the relative cost differences between type I and type II errors. To interpret model performance meaningfully, it is necessary to differentiate between these costs. However, as Koh (1992) pointed out, the costs of incorrect classifications are largely intangible and not easily measured. In addition, different stakeholders in a company may not treat type I and type II error costs equally in terms of risk and seriousness. Much debate has surrounded this issue in the corporate failure literature.

From a creditors point of view, a type I error of misclassifying a failed firm will probably be more costly than a type II error of misclassifying a non-failed firm (Beaver, 1966; Lau, 1987). However, Blum (1974) pointed out that a type II error may be more costly in certain situations. For example, in the antitrust decision, allowing a non failed company to complete a merger because of their mistaken description as a failing firm may do more injury to the interests of the public than sending a genuinely failing firm to its liquidation.

Costs of errors also depend on the industry under investigation. In the banking context, Looney et al (1989) estimated that the cost of a type I error (misclassification of a failing bank as sound) to banking authorities is roughly 58 times more costly than a type II error (misclassification of a healthy bank as failing). Altman (1977) estimated the relative costs of type I to type II errors for commercial bank lending as being 7:1. Type II errors, however, involve unnecessary examinations and a loss of public confidence in the affected banks. This may result in a run on the bank.

Ultimately, any weights assigned to type I and type II error costs are subjective in nature (Etheridge & Sriram, 1997, 242).

10.1.4. MEASURING ERROR RATES

Type I and type II error rates can be interpreted as the probability of error conditional on either:

- the actual status of the firm (for example, the number of type I errors divided by the number of actual failures in the sample)
- the prediction made (for example, the number of type I errors divided by the number predicted failures).

However, Beaver (1966, 88) noted that if the probability of failure for the sample differs from that of the total population, an inference of percentage of total errors to the population is not meaningful.

When using many of the classical statistical classification techniques, the number of predicted bankruptcies will depend on a selected cut-off point (Lennox, 1999). For example, in a logit analysis, if the cut-off is set equal to 0.1, a company for which the expected probability of bankruptcy exceeds 10% will be forecast as an impending bankruptcy. However, a company for which the expected probability of bankruptcy is less than 10% is predicted to survive. Lennox noted that one can simply increase the number of companies that the model predicts as impending failures by reducing the cut-off probability for the logit, probit and MDA models. This creates a trade-off between type I and type II errors. The objective is to minimise these errors while bearing the cost of misclassification of each error type in mind.

A perfectly accurate model would classify 100% of the firms correctly. However, consider a hypothetical situation in which the two sample groups (failed and non-failed) to be classified are significantly different in size. A situation may arise in which the proportion of the large group (non-failure) is classified almost completely correctly, while only a relatively small proportion of the smaller, but more crucial "failure" group, is accurately identified (Korobow and Stuhr, 1984, 269). In this case, "percent classified correctly" is not an informative measure. For this reason, Korobow and Stuhr devised a "weighted efficiency" (WE) measure of model performance:

$$WE = \frac{BWF}{VB} * \frac{BWF}{TWF} * CC \qquad (10.1.)$$

Where:

- CC = percent classified correctly,
- BWF = failures correctly identified by the model,
- VB = all failures identified by the model,
- TWF = total number of failures in the sample.

WE is sensitive to both the percentage of failed firms classified correctly and the percentage of those firms identified as failures by the model.

Korobow and Stuhr found the WE measure to be sensitive to a trade-off between the percentage of failures correctly identified and the percentage of the sample that the model identified as "failed".

10.1.5. Analysis of Misclassifications

Looney et al (1989) analysed the banks that were misclassified in their study in terms of their geographic location. They made interesting observations with regards to the relationship between the misclassifications and the economic situations and legal frameworks in different states. The analysis enabled them to identify causes that were symptomatic of failure; these included a weakened local economy and situations in which agricultural loan losses were sustained.

Looney et al then attempted to identify which ratios most significantly contributed to the misclassifications. Each of the misclassified banks' ratios were normalised (i.e. by subtracting the sample ratio average and then dividing through by the sample ratio standard deviation). They counted the number of observations with the "wrong" sign for each model and each ratio in the model. For example, if a bank failed, one would expect that the return on assets for that bank would be lower than the sample average. If a type I error occurred and a failing bank was incorrectly classified as healthy, then a "wrong" sign would be one that indicated a return on assets greater than that of the sample average. A ratio associated with a large proportion of misclassifications was identified as relating to a potential cause of the misclassification. In the study, a ratio was deemed "important" if it exhibited the "wrong" sign for more than 50% of the observations.

Analysis of the errors used in calculating the error rates for a model has not been performed regularly in the corporate failure literature. This study does perform such an analysis.

10.2. TESTING THE SIGNIFICANCE OF THE MODEL

Instead of evaluating the results of a model, one can test the attributes of the actual model for significance. Each classification technique has different parameters and various scoring systems. The following are some of the more common evaluation methods that can be used to assess such techniques.

10.2.1. F-RATIO

The F-value can be used to test the overall discriminating power of a model. This value is the ratio of the sums-of-square between-groups to the within-groups sums-of-square. The ratio is calculated as follows:

$$F = \frac{\sum_{g=1}^{G} N_g(\bar{y}_g - \bar{y})^2}{\sum_{g=1}^{G}\sum_{p=1}^{N_g}(y_{pg} - \bar{y}_g)^2} \qquad (10.2.)$$

Where:

- G = Number of groups (usually two)
- g = Group number
- N_g = Number of firms in group g
- y_{pg} = Firm p in group g, p = 1... N_g
- $\bar{y}_g$ = Group mean (centroid)
- $\bar{y}$ = Overall sample mean

This ratio is maximised in a situation in which the means of the *G* groups are spread further apart and, simultaneously, the dispersion of the data points (for each individual firm) within each group is smaller. Phrased differently, the F-ratio tests how well a model discriminates between groups and how well defined each individual group is (Altman, 1968, 398).

An implicit assumption of such a test is that the model must produce some value or score (y) that can be used to discriminate between groups. Thus, the F-ratio can be used in the evaluation of the multiple discriminant analysis model.

10.2.2. Likelihood Ratio Test

The likelihood ratio test (Zavgren, 1985) provides a more definitive test of strength for the logit and probit models. It begins by using the null hypothesis that the entire model is insignificant, i.e. that all the coefficients are insignificantly different from zero. The likelihood ratio is defined as:

$$\lambda = \frac{L(\vartheta^*)}{L(\vartheta)} \tag{10.3.}$$

Where

- $L(\vartheta^*) =$ the value of the likelihood function estimates for the unrestricted model
- $L(\vartheta) =$ the value of the likelihood function with the restriction imposed that all coefficients are zero.

The test statistic of *-2logΛ*, which is distributed in a Chi-square with degrees of freedom equal to the number of independent restrictions imposed, can be tested for significance.

10.2.3. Shannon's Entropy Theory

Shannon's Entropy Theory was applied by Zavgren (1985, 30) to his probabilistic financial failure model. Seen in an information theory context, the probability estimates generated by the logit model are messages from an information system. The quantity of information in each message can be measured by its ability to reduce uncertainty. "Entropy" is defined as the degree of uncertainty over the occurrence of an event. A more extensive description of this theory is discussed in the study performed by Zavgren.

CHAPTER 11
IN SUMMARY: REVIEW OF EMPIRICAL RESEARCH ON CORPORATE FAILURE PREDICTION

From the literature reviewed in this section it should be clear that there has been extensive empirical research in the field of corporate failure prediction over the last four decades. Despite the attention that this subject has drawn, there is very little conclusive evidence as to what constitutes a single optimal failure prediction model. There are many subjective areas within each step of the process described above.

However, because of the importance of this topic to the social and economic environment, and because of the diversity and extent of different stakeholders who are impacted by failing firms, this area of research continues to attract researchers keen on applying potentially superior methodologies to the problem.

The following sections seek to report the application of a new methodology, not applied in any research reviewed by this study, to the problem in a South African context. While many aspects of this construction process remain subjective in nature, the research that follows draws on the findings reviewed in Section A in order to justify this study's approach to corporate failure prediction.

CHAPTER 12
INTRODUCTION: MODEL CONSTRUCTION

The steps followed in the construction of a corporate failure prediction model are materially homogeneous across studies performed in different geographic locations, over different time periods and using varying methodologies. Studies performed in South Africa are of no exception.

The following sections of the report lay out the research performed in constructing an empirical corporate failure prediction model in South Africa using a machine learning approach. All data was captured in Microsoft Excel before being imported into Matlab 6 Release 12, Student Version.

The report follows the basic steps for the construction process that are laid out in Section A. In summary, the following sections continue as follows:

Section B: Sample and Data Preparation

- **Definition of Corporate Failure in this Study**
  The next chapter defines the concept of corporate failure as it is to be used in this study.

- **Data Collection**
  Chapter Fourteen discusses the failed and non-failed company sample selection process, as well as the collection of data related to these samples. This step in the model construction process has not been well documented in other corporate failure studies. However, this study presents a detailed discussion of the data sources, sample selection procedures and limitations and a brief analysis of the final samples.

- **Feature Selection**
  Chapter Fifteen presents a discussion of the uses, advantages and limitations of financial statement information and ratio analysis before proceeding onto a description of the broad groups of ratios and data selected for input into the feature subset selection process. The calculations and justifications for each data item are presented as part of this chapter.

- **Data Pre-Processing**

  Chapter Sixteen discusses how the features chosen in Chapter Fifteen are processed so as to allow for the construction of an optimal classifier using the machine learning approaches adopted in this study.

Section C: Feature Subset Selection

Chapter Seventeen discusses the mechanics of and prior research related to Population-Based Incremental Learning (PBIL) and Thornton's Separability Index (SI). These techniques are then applied to the problem of feature subset selection in Chapter Eighteen.

Section D: Classifier Construction and Evaluation

- **Classifier Construction**

  Two methodologies are then applied to the construction of the classifier. Firstly, Chapter Nineteen uses a k-Nearest Neighbour (kNN) approach. Then Chapter Twenty applies Kernel Ridge Regression (KRR) to the problem. In addition, each chapter contains a discussion of the mechanics, prior research and justifications for the application of the respective technique. In particular, the discussion of KRR in Chapter Twenty seeks to justify this technique as a state-of-the-art method applicable to this research problem.

- **Evaluation of Models**

  The final chapter evaluates the different models constructed based on various error cost ratio assumptions.

The end product of this report is a number of failure prediction models:

- A model is presented for predicting failure one, two and three financial year ends in advance;
- For each forecast period, there are a number of optimal feature subsets selected and applied to failure prediction;
- For each feature subset within each forecast period both the kNN and KRR techniques are applied.

The research performed in reviewing the corporate failure literature in Section A is drawn on in each step of the model construction process. In addition, where new methodologies are applied, prior studies are reviewed with a view to explaining and justifying the methods as they are applied in this study.

CHAPTER 13
DEFINITION OF CORPORATE FAILURE IN THIS STUDY

As discussed in Chapter Three, there are varying degrees of failure. With this in mind, a "spectrum of corporate failure" was defined. This spectrum ranges from the complete demise of bankruptcy to failure in terms of falling short of a required rate of return on invested capital.

This chapter defines corporate failure as it has been used in this study. The discussion begins by looking at the aforementioned spectrum. The failed company criteria are then distilled from the chosen definition. The definition of failure and the resulting criteria are then justified in the context of a South African study and in terms of the data availability for this particular study.

13.1. SCOPE OF DEFINITION IN THIS STUDY

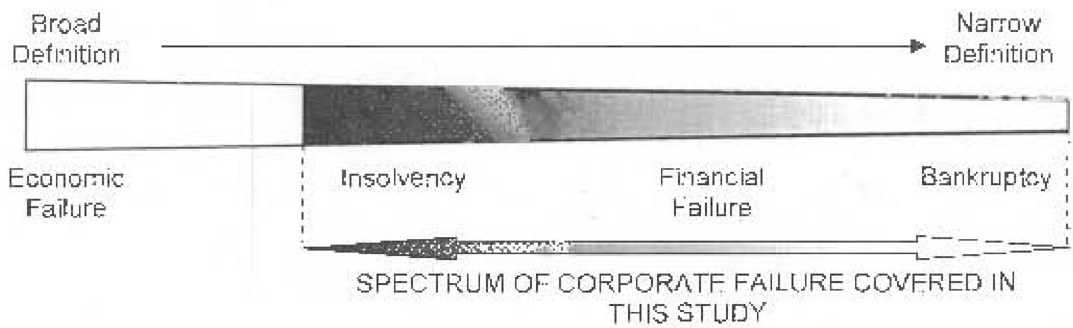


Figure 13.1. Spectrum of Corporate Failure: Definition for this study

As is the case with the majority of South African corporate failure studies, a broad definition of failure was assumed in this study. The illustration above indicates the range of the spectrum, as defined in Figure 3.1., which this study has encompassed.

A broader definition means that failure in this study has included situations in which the failed company:

- has not been able to meet current financial obligations, or

- has not been able to operate profitably

and, as a result, has ceased existing as the same ongoing entity.

A narrow definition of failure was not used for the following reasons:

- In South Africa, it is difficult to isolate corporate bankruptcies. This is because the liquidation process in South Africa can be triggered for many non-failure related reasons. This is discussed extensively in Chapter Three (see subsection 3.3.1.). As a result it was not feasible to restrict the definition to bankruptcy alone.
- There are a small number of companies listed on the JSE Securities Exchange, the population from which the sample for this study was drawn, relative to those companies listed on the exchanges of, for example, the U.S. As a result, the number of companies that have failed in South Africa during the period under investigation based on a narrow definition of failure is too small to allow for a significant sample size. An expanded definition allowed for the possibility of a larger sample.
- The model in this study was constructed to predict failure in terms of poor financial performance and the inability of a company to meet its current financial commitments. Conceptually, many stakeholders will be interested in such a scenario.

13.2. FAILED COMPANY SELECTION CRITERIA

A number of possible symptoms may manifest themselves when a firm fails. It is necessary to identify what, specifically, these may be in order to distil from the definition of failure a set of criteria that can be used for selecting a sample of failed companies.

The symptoms used as criteria for failed company selection in this study are presented below. The following subsection then presents an explanation of and justification for these selection criteria and compares them to those used in other South African corporate failure studies.

Note that the population of listed companies from which the sample of failed companies was drawn included only those companies that had been listed on the JSE Securities Exchange for at least two years. The reasons for the exclusion of companies listed for only a single year are discussed in the following chapter.

13.2.1. FAILURE CRITERIA FOR LISTED COMPANIES WITH AT LEAST THREE YEARS OF DATA PRIOR TO DATE OF FAILURE

A company listed on the JSE Securities Exchange for at least three years prior to it being:

(A) delisted or suspended from being traded on the JSE, or
(B) taken over by another company, or
(C) recapitalised through the issue of a significant value of shares, or
(D) converted into a cash shell prior to one of the above corporate events;

is deemed to have failed if at least one of the following criteria apply:

(1) The company has not declared an ordinary dividend for two out of the three years preceding the date of failure.
(2) The company has made negligible or negative net profits for two out of the three years preceding the date of failure.
(3) The company has a negative net worth in the last financial statements audited prior to the date of failure.
(4) The company has current liabilities that exceed current assets for two of the three years preceding the date of failure.
(5) The market capitalisation of the company has fallen by 60% or more from the financial year end three years prior to failure until the financial year end immediately preceding the date of failure.

For the purposes of this study:

- **The date of failure is defined as**:
  the date on which one of the corporate actions, A to D above, occurred. In other words: (A) the date on which the company was delisted/suspended from the JSE, or (B) the date on which the company was taken over by another company, or (C) the date on which a significant value of shares were issued in order to recapitalised the company, or, finally, (D) the date on which the company became a cash shell prior to one of the events (A) to (C).

- **Recapitalisation is defined as**:
  the situation in which a company increased its issued share capital by more than 80% through a share issue.

- **A cash shell is defined as**:

a company which has disposed of almost all of its fixed assets and in which the working capital consists almost solely of cash.

13.2.2. Failure Criteria for Listed Companies with only Two Years of Data Prior to Date of Failure

A company listed on the JSE Securities Exchange for two years prior to it being:

(A) delisted or suspended from being traded on the JSE, or
(B) taken over by another company, or
(C) recapitalised through the issue of a significant value of shares, or
(D) converted into a cash shell prior to one of the above corporate events;

is deemed to have failed if at least one of the following criteria apply:

(1) The company has not declared an ordinary dividend for both years preceding the date of failure.
(2) The company has made negligible or negative net profits for both years preceding the date of failure.
(3) The company has a negative net worth in the last financial statements audited prior to the date of failure.
(4) The company has current liabilities that exceed current assets for both years preceding the date of failure.
(5) The market capitalisation of the company has fallen by 60% or more from the financial year end two years prior to failure until the financial year end immediately preceding the date of failure.

The date of failure, recapitalisation and a cash shell are defined as described in subsection 13.3.1.

13.2.3. Reference to Reason for Failure

Throughout the remainder of this study, the reasons for failure will be referred to as follows:

- The letters A to D (corresponding with the above lettering) will refer to the corporate action that resulted in the firm being identified as a failure.
- The numbers 1 to 5 will refer to the failure criteria corresponding to the criteria listed above.

For example, a "B3" failure will refer to a company that has been taken over after it was identified as having had a negative net worth in the year preceding failure.

13.3. JUSTIFICATION FOR FAILED COMPANY SELECTION CRITERIA

Each company that was defined as having failed per this study needs to have cleared two hurdles:

- Firstly, the company has to have been subject to one of the four "corporate actions/scenarios" (listed A to D above).
- Then the company that has been subject to one of these corporate actions needs to have met one of the "poor performance criteria" (listed 1 to 5 above).

13.3.1. Discussion of Corporate Actions

The corporate actions, listed as A to D above, seek to identify those situations in which a company ceases to exist in its original operational or marketable form, or the situation in which it significantly changes its capital structure.

(a) Delisting or suspension from the JSE Securities Exchange

A company that is delisted or suspended from the JSE Securities Exchange is a significantly less marketable investment.

(b) Company takeover

Similarly, when a company is taken over by another, it can no longer be marketed as a separate entity and, in most cases, its operations will be changed to synergise with those of the acquiring company.

(c) Recapitalisation

When a company is in financial distress it may, under certain circumstances, decide to issue a large number of shares in order to raise funds so as to allow the business to settle debts and fund operations. In such situations, the capital structure of the going concern after recapitalisation is significantly different to the structure in place prior to such a share issue.

The recapitalisation of a company is usually a step that the management and shareholders of a company will want to avoid. The issue of additional shares dilutes the control of existing shareholders. In addition, recapitalisation in times of distress will result in the additional shares being issued at a relatively low price, resulting in a dilution of control without maximising fund inflows. For such reasons, the recapitalisation of a company is a final step that would usually only be taken as a last resort.

(d) *Cash shell*

A cash shell is defined above. This situation may arise when the operations of a company are winding down and, as a result, the income producing assets are sold off and all working capital balances are run down to nil balances. The main asset that remains in the business at such a point will be a cash balance.

Such a curtailment signals a significant change in the operations of a company.

13.3.2. DISCUSSION OF "POOR PERFORMANCE CRITERIA"

The delisting, takeover, recapitalisation or conversion to a cash shell of a business, when considered alone, is not necessarily an indication that a company has failed. For example:

- A company may be delisted as a result of a downsizing of the business or in a situation in which the costs related to the listing outweigh the benefits to the company.
- A company may be taken over by a competitor as a result of the target's excellent performance or in order for the acquirer to acquire any significant cash reserves held by the target.
- A company may be recapitalised in order to fund such a takeover or in order to significantly increase the scale of its operations.
- A cash shell may be created as a result of a company's profitable operations being transferred into another entity for reasons including economies of scale.

However, when these drastic changes in the existence of a company are coupled with an indicator of poor financial performance and/or position, these corporate events are deemed to be evidence of corporate failure, and vice versa. In other words, one can view zero dividend declaration, poor profitability, factual and potential commercial insolvency as individually non-critical or temporary in nature, unless they occur in combination with a drastic change in the existence of a company. At such a point, the combined criteria can be deemed to be an indication of the failure of that company.

Specifically, these indicators were selected so as to identify those companies where poor profitability and the inability to meet current financial obligations resulted in one of the corporate actions listed above. In this way, the criteria for failed company selection are consistent with the scope of corporate failure adopted in this study.

(a) Dividend policy

Modigliani & Miller (1961) argued that, in a perfect world, the value of a firm is unaffected by the dividend policy or a change therein. However, research has consistently shown that stock prices in fact move in the direction of a change in dividend. The signalling arguments developed by Bhattacharya (1979) and John & Williams (1985) present the basis for arguments of asymmetric information between managers and shareholders.

Given this environment, negative information will be withheld until financial constraints force the release of such information (Ryan, Besley & Lee, 2000, 35). Therefore, the payment of a dividend, although not compulsory, can be deemed to be a financial obligation on a company because of the signal that non-payment may send to its investors.

A single year of dividend non-payment may still be as a result of non-failure related reasons. However, two years of dividend non-payment, coupled with one of the aforementioned corporate actions, can be deemed to be as a result of the company being unable to its meet current financial obligations. For this reason, companies that were identified as not having paid dividends in two out of the three years prior to one of the corporate actions, A to D, were classified as failures.

(b) Net profit

Any company can go through a lean year in which accounting profits are negligible or even negative. However, where such a scenario occurred for two out of three years prior to a takeover or such event, the poor profitability of the company was interpreted, *ex post*, as a visible symptom of the failure process.

(c) Net asset value

A negative net worth is a direct measure of company insolvency. Where such a situation arose in the year preceding one of the corporate events A to D, the negative net worth was deemed to have contributed to the occurrence of the said corporate event. This symptom is a measure of a narrower failure definition encompassed within the broader definition used in this study.

(d) Net current assets

Net current assets are a direct way of measuring the ability of a company to meet current financial obligations. Negative net current assets for two out of three years prior to one of the corporate actions A to D were deemed to be an indication of a consistent cash flow problem that resulted in the failure of the company.

(e) Market value

The efficient market hypothesis suggests that the market should discount poor financial performance and potential failure into a share price. If a company is in the process of failing then one would expect the share price to fall. If a company has already failed in the market, the share price of that company will not fall much further, and that company will not be selected under the last criterion. This is the desired effect, as the model that is built in this study attempts to predict future failures, and not failures that have already happened.

A fall in absolute return, rather than a return in excess of the market, was used to measure failure. In times of a bear market, all company market values may be falling in absolute terms. However, these decreases in market value do not all result in failure. The view was taken in this study that an absolute fall in value, coupled with a termination in the existence of a company, is an indication that the company has failed. This may have been because the said company was unable to weather the downturn in the market.

13.4. COMPARISON TO PRIOR SOUTH AFRICAN RESEARCH

The criteria used in other South African corporate failure studies have been summarised in Table 3.1. and appear in bold below. The criteria used in this study were selected with these in mind:

- **Company liquidated or placed under judicial management**:
  For the reasons mentioned in Chapter Three, this alone is not an indication of failure. However, these corporate scenarios have been captured within the scope of A in the selection criteria.

- **Company had a negative net worth**:
  This has been incorporated into this study.

- **Company reduced share capital to bring it in line with related assets**:
  This is an indication that the company has experienced a sustained period of poor profitability. This symptom of failure has been addressed by the negative net worth and net profit criteria of this study.

- **Company failed to pay preference dividend**:
  A company is obligated to pay preference dividends before ordinary dividends. As such, by having included the non-payment of ordinary dividends as a criterion in this study, a broader scope of failure was adopted. Preference dividend payment was, however, used as a potential feature input for the prediction model constructed.

- **Company had poor financial performance**:
  The criterion for poor profitability used in this study is a combination of the criterion used by Daya (1977) and Immelman (1980). Daya combined poor profitability with subsequent delisting while Immelman set a hurdle of two years of negative net profits.

- **Company failed to pay ordinary dividend**:
  A similar criterion as that used by Le Roux (1980) has been adopted for this study.

- **Takeover of company coupled with another criteria:**
  This has been used in this study.

13.5. DEFINING "NON-FAILURE"

Finally, a company was defined as "non-failed" if it was still listed, and not suspended, as at 30 June 2003 and if it was not recapitalised or taken over during the period for which it was selected.

CHAPTER 14
SAMPLE SELECTION AND DATA COLLECTION

The failure criteria (1 to 5) defined in Chapter Thirteen were then applied to the population of companies identified as having undergone one of the corporate actions, A to D, between 1 January 1996 and 30 June 2001. This chapter describes this failure selection process, the process for the selection of the non-failed portion of the sample and the source and type of data collected for all companies in the final sample. The chapter concludes with a brief analysis of the characteristics of the final sample.

14.1. DATA SOURCES

There are numerous financial information service providers that collate and publish financial statement data for South African listed companies. Amongst these, BFA McGregor collates and publishes such data in both electronic and printed format. The BFA McGregor service includes a database of share price and market index movements for the JSE. In addition, the JSE periodically publishes a handbook containing summarised listed company financial and non-financial information.

14.1.1. AUDITED FINANCIAL STATEMENT INFORMATION

The financial statement data used in this study was obtained from the "Standardised Financial Statement Database" on the BFA McGregor RAID Station service. This database captures all financial statement information in a format that is standardised across all listed companies and years contained within the database (for the format of the standardised information, see Appendix C).

This single source was used for all financial statement-related information in order to assure consistency in the type and calculation of data used to build the failure prediction model.

The financial information contained within this database is widely utilised by financial and academic institutions in South Africa. Therefore, the additional advantages in using this data source include:

- easy reperformance of the work in this study, and
- an enhanced applicability of the model constructed for use in a commercial setting.

14.1.2. JSE Market Data

A database containing JSE index and share price returns is available on BFA RAID Station. This database was used as the source for such market-related data.

14.1.3. Other Non-Financial Data

Additional non-financial data, not available from BFA RAID Station, was sourced from the JSE Handbook. This data included:

- Date of commencement of operations of each company
- Date of listing of each company
- Sector and nature of business of each company

The BFA McGregor handbook was consulted for data regarding cross-holdings within group structures and for information relating to the reasons and dates for the delisting or suspension of companies from the JSE.

14.2. FAILED COMPANY DATA COLLECTION

14.2.1. Data on Delisted and Suspended Companies

Historically, BFA RAID Station has removed the financial statement data for a company from the database when that company was delisted. However, for delistings after 1996, this data was transferred to a separate category on RAID Station specifically assigned to delistings. Standardised balance sheets, income statements, cash flow statements, value-added statements and additional sundry information (as per Appendix C) for all companies contained under the "delisted" category on RAID Station were downloaded into Microsoft Excel format. The BFA McGregor handbooks for the years between 1996 and 2003 were consulted to obtain the date and reason for delisting for each company.

In addition, the BFA McGregor Handbooks for 2002 and 2003 were consulted for the names of companies suspended from the JSE but not yet delisted. The data for such companies was downloaded from BFA RAID Station.

In this way, audited financial data was obtained for each company for the period up to the last date prior to delisting or suspension for which the company had published such data.

Companies were then deleted off the downloaded Excel database if:

- the company did not have at least two years of data available, or
- the financial data of the company was not available in Rands, or
- the company had operated in the financial services or mining sectors.

Companies with only a single year of financial data were not included in this study. Companies that were only listed for a year before failure were considered to have insufficient available information to be included in the model. This is because it is not possible to compare ratios over a period of time for such cases.

In order to control for any differences in exchange rates over the period of study, ratios were only calculated from financial data that was denominated in Rands.

Financial services and mining companies were excluded from the sample as these industries are significantly different in terms of their structure and their financial accounting framework. A few of the key differences can be highlighted as follows:

- **Financial services companies**:
  There is no trading of physical goods in this industry. Thus, the investment in fixed and current assets used for trading purposes is non-existent. In addition, there is no sales revenue, a key financial indicator used in building models for retail/wholesale/manufacturing companies. The comparable key indicators in the financial services industry would be interest income or net interest income for banks, or premiums for insurance companies. However, the fluctuations in these revenues are driven by completely different micro- and macroeconomic variables than those that drive retail sales.

- **Mining companies**:
  Mines are depleting assets. In South Africa, mines have used an alternative form of accounting suited to the historical structure of mining operations, known as the appropriation method. Subsequently, with mining companies having restructured to allow for operations to continue perpetually, there has been a move towards standard GAAP. This change makes comparison across companies and years difficult.

Other South African corporate failure studies have controlled for industry in a similar manner (Daya, 1977; Court, 1992; Arron, 1994; Chapter Four for further detailed discussion). Ohlson (1980) went on to identify utilities and transportation companies as structurally different from industrial companies. But as Platt et al (1984) noted, often there are an insufficient number of publicly traded companies to segment the analysis further than the manufacturing/retail/service companies that have been used in most studies.

14.2.2. Failed Company Selection

The following steps were followed in order to select the final sample of failed companies for this study:

- All data was inspected to determine if any company on the downloaded Excel database met the definition of a cash shell as defined in the previous chapter. Any years of company financial information meeting this definition were deleted from the database.

- The percentage increase in the number of issued shares for each year was calculated across all companies in the database after taking into account any share splits. The company was deemed to have recapitalised itself if this increase was greater than 80% (as defined in Chapter Thirteen).

  This study used financial data for the three years preceding the date of failure. If the recapitalisation occurred within the three years of financial statement data preceding delisting/suspension, the date of potential failure was set in the year of recapitalisation rather than on the date of delisting/suspension. The three years of data preceding the recapitalisation was then utilised for model construction.

  This resulted in a number of companies being entirely removed from the database in situations in which less than two years of data remained after this elimination process.

- The remaining companies on the Excel database were then tested to determine whether they had failed as defined in Chapter Thirteen, or whether they had been delisted, suspended or recapitalised for reasons other than failure.

  These tests were run across the two or three years of data (depending on availability) prior to the delisting, suspension, recapitalisation or cessation of operations of the company. All companies that did not meet at least one of the failure criteria as defined in Chapter Thirteen were then deleted off the Excel database.

- The BFA McGregor Handbook was then consulted to determine if any significant cross-holdings existed between the companies that remained in the sample of failures. If a subsidiary and its holding company were both included in the sample, the failure information relating to those companies would effectively have been double-counted in the construction and testing of the failure prediction model. For example, Stocks & Stocks Ltd was removed from the database while its holding company, Stocks & Stocks Holdings Ltd, remained in the sample of failures selected.

- Finally, two investment professionals that work within the equity financial investment services industry in South Africa were consulted. These professionals inspected the final list of failures to determine whether, to their best knowledge, any of the companies on the list had not failed or were companies held within the same group.

The remaining database was then trimmed to only include a maximum of three years of data prior to the delisting, suspension or recapitalisation of each company.

14.2.3. DEFINING THE DATE OF FAILURE

It is critical to have the failure date defined in order to correctly identify the pre-failure financial data. If the failure date is not accurately defined, for example if post-failure data is mixed in with pre-failure data, then the predictive model constructed from the data could appear to have better predictive power than is really the case (Bortiz, Kennedy & Alberquerque, 1995, 100).

In this study, financial data was only used for years of operation prior to the delisting, suspension or recapitalisation of the company. In this way, the use post-failure data was avoided.

14.2.4. ANOMALY TEST ON FAILED COMPANY DATA

The last date on which shares were traded for each failed company was looked up on BFA RAID Station. A test was then run on the database of failures to confirm that such date was after the last date for which financial data for that company was utilised in constructing the failure prediction model.

14.3. NON-FAILED COMPANY DATA COLLECTION

14.3.1. METHOD OF NON-FAILED COMPANY SELECTION

> "If the comparison of failed and non-failed firms is to be meaningful, the sample of non-failed firms should be drawn from the same population as that of the failed firms." (Beaver, 1966, 76)

A paired sample technique was employed in the selection of non-failed companies. Beaver (1966, 73) justified the paired approach by arguing that it is necessary to provide "control" over factors that otherwise might blur the relationship between ratios and failure. Most South African studies have employed this approach.

Non-failed companies were selected using a pairing process based on the following criteria:

- **Industry sector**:
  Beaver (1966, 73) noted that the same numerical value of a ratio across different sectors may imply a different probability of failure. While the sample of failed companies was insufficient to segment the analysis of companies into sectors, pairing failed companies to non-failed companies in the same sector allowed for the control of sector-influence to some extent.

- **Year of failure**:
  Controlling for the year of failure is important as, Platt et al (1994) noted, time series distortions may result when financial data is spread over different stages of the business cycle and across varying economic conditions.

- **Company size**:
  Company size was measured using average total assets (excluding intangibles) and turnover. Total assets exclude intangibles as the value of intangibles can rarely be measured accurately. Beaver suggested that the ratios of firms from significantly different asset-size classes cannot be directly compared (1966, 75). In addition, it has been shown that smaller companies have an inherently greater risk of failure (Alexander, 1949). A randomly selected sample of non-failed companies would, thus, have been of a greater average size. This would have made any direct inferences to the population difficult.

It should be noted that while a paired sample design mitigates the disruptive influence of the aforementioned factors, it also virtually eliminates any predictive power these factors may have had. A complete discussion of the paired and non-paired approaches to sample selection can be found in Chapter Four.

14.3.2. PAIRED SAMPLE SELECTION PROCESS

The following steps were followed in the pairing process:

(a) *Industry sector*

Using the BFA McGregor Blink Server, all companies listed on the JSE as at 30 June 2003 were downloaded. The download included each company's JSE code, short name, full registered name, sector, sub-sector and the date of the most recently released audited financial statements included on the BFA RAID Station database.

To quote the Blink Server manual, this server allows for the direct integration of the BFA RAID Station database into Excel spreadsheets.

As the RAID Station database does not store the sector for companies delisted from the JSE, the sector for each failed company was looked up in the relevant year of the BFA McGregor Handbook.

For failed and listed companies operating in the venture capital market (VCM) and development capital market (DCM) sectors, an additional industry sector was assigned based on the nature of the business as described in the BFA McGregor Handbook. This was in order to maximise the accuracy of pairing based on the nature of each company's business.

The sample of listed companies that were used to find a match for each failed company was based on an initial pairing of these sectors. Failed companies operating in the VCM or DCM sectors were initially paired based on nature of business and then based on VCM/DCM wherever possible. In this way, the study controlled for the additional risk factors relating to companies operating in the VCM/DCM sectors.

(b) *Year of failure and data availability*

Data for the non-failed listed sub-sample (determined through matching based on sector) were then matched to failed companies based on the year of failure.

Non-failed companies that were not listed at least three years prior to the year of failure of the failed company to which it may have been paired were not considered. This was because the model to be constructed needed three years of data.

(c) *Average total assets (excluding intangibles)*

Using the Blink Server, the total assets (excluding intangibles) and turnover for all years from 1993 to 2003 for every company listed on the JSE at 30 June 2003 were downloaded from the RAID Station standardised financial statements.

The average total assets (excluding intangibles) was then calculated for the period covering the three/two years for which data were available for the potential failed company match.

A non-failed company was then selected by matching this three year total assets average to that of a failed company. This non-failed match was drawn from the sub-sample of non-failed companies that matched the sector and year of failure (as determined in the previous two steps).

(d) *Average turnover*

If there was no non-failed company that closely matched the failed company based on total assets, turnover was used in a process identical to the one described in *(c)* above.

14.3.3. DATA ON NON-FAILED COMPANIES

After the non-failed companies had been selected using the process described above, the relevant financial statement data were downloaded from BFA RAID Station. Companies were eliminated if financial data were not available in Rands. New non-failed companies were then selected.

14.3.4. ANOMALY TEST ON NON-FAILED COMPANY DATA

In order to ensure that no failed companies were incorrectly identified as non-failed, the following procedures were performed:

- A company was only selected as non-failed if it did not appear on the list of delisted and suspended companies published in the BFA McGregor Handbook.

- The companies were tested for recapitalisation as defined in Chapter Thirteen and excluded if recapitalised.

14.4. RESULTS OF THE PAIRING PROCESS

14.4.1. Final Sample Selected

The final sample of failed and non-failed companies is included in Appendix D. Each of the 82 failed companies has been assigned a reference code of F1 through F82 for ease of reference. The non-failed pairings are numbered N1 through N82 and have a number corresponding to their failed match.

The appendix also sets out which failure criteria have been met in identifying the failed companies, as well as the year of failure, sector and the number of years of available data for each company.

Of the sample of failed companies selected, 61 companies (74%) had three years of data available prior to failure. The remaining 21 companies (26%) had only two years of data available.

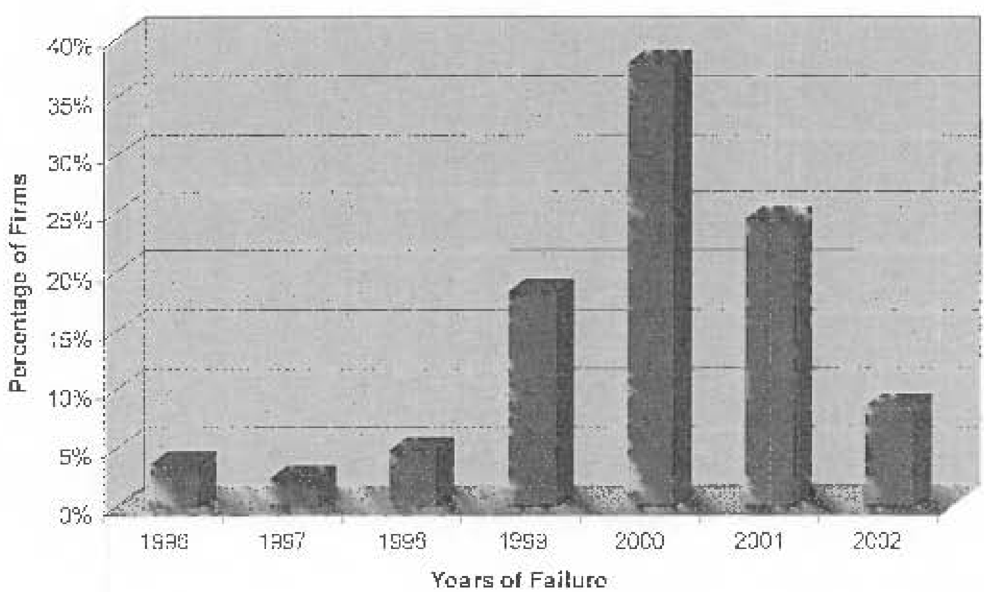


Figure 14.1. Distribution of year of failure of selected failed companies over the period under study.

The distribution of failures across the period under study (1996-2002) was heavily weighted towards 2000 and 2001, with 62% of the sample falling within these two years. The four years from 1999 to 2002 held 89% of all the sampled failed companies. The small sample selected from 1996 to 1998 is as a result of the prior policy of BFA McGregor RAID Station to eliminate any company data when that company was delisted. As a result, the database is not complete as it was only updated going back to 1996 in 1999.

The heavy weighting of failures over the period 1999 to 2002 was considered to be a positive contributing factor to the model for the following reasons:

- The time series distortions from changing economic conditions are reduced when a shorter time period is under study.
- As the period is recent, the model is more applicable at the time of its completion.

14.4.2. Accuracy of Pairing Process

(a) Pairing based on sector

Of the sample of the 82 failed and the 82 non-failed companies selected, 78 pairs were matched by sector. There were an insufficient number of non-failed listed companies in the telecommunications sector available for matching purposes. As a result, four companies (PARADIGM, PRADTECH, RADIOSPR and VALUECOM) were matched with companies in the information technology (IT) and electronic & electrical equipment (EE) sectors. The IT and EE sectors were deemed to be the best alternatives to the telecommunications sector because of the similarity between these sectors with regards to the rate of technological development and the types of services provided and products produced.

There are 15 failed companies from either the venture capital market or development capital market in the final sample. These companies have been marked as VCM or DCM in the "Sector" column of Appendix D. Of these, only four companies were matched to companies also operating in the VCM or DCM. This is due to the fact that the sample of failed VCM/DCM companies was significantly larger than the sample of companies that was still listed in these sectors as at 30 June 2003.

(b) Pairing based on size

The table below indicates how closely matched the non-failed sample is to the failed sample based on size.

Range above and below failed sample	Proportion of non-failed companies within range
±10%	17%
±20%	38%
±30%	48%
±40%	57%
±50%	62%
±60%	64%
±70%	68%

Table 14.1. The percentage of non-failed companies falling within a range of *x* percent above and below either the average total assets or turnover of the failed company with which it was paired.

In order to illustrate the differences in the sizes of the selected failed and non-failed samples, the means and standard deviations of the assets and turnover of these samples are presented below.

Amounts in R `000 000	Non-Failed Companies	Failed Companies	Percentage Difference
3 Yr Average Total Assets			
Mean	R 561	R 390	-30.48%
Standard Deviation	R 1 107	R 987	-10.84%
3 Yr Average Turnover			
Mean	R 750	R 491	-34.53%
Standard Deviation	R 1 255	R 1 009	-19.60%

Table 14.2. Differences between the 3 years average of total assets and turnover across the failed and non-failed samples.

It is clear from the above statistics that the mean average total assets and turnover for the failed sample is about a third smaller than that of the non-failed sample. This is consistent with the findings of corporate failure studies which show that smaller companies have an inherently greater risk of failure. The means of the two samples could not have been matched any closer due to the unavailability of surviving smaller companies listed on the JSE Securities Exchange as at 30 June 2003.

The larger standard deviation of the non-failed sample is as a result of larger companies having been used as matches in situations where there were an insufficient number of small companies on which to draw.

The outcome of the sample selection process is a sample matched, as closely as possible, based on size after having taken the other selection criteria into account.

14.4.3. Analysis of the Reasons for Failure

The table below identifies the most commonly met failure criterion within the failed company sample - a cessation of dividend payments (88% of failed firms met this criterion). A falling market value and negative profit were the next most frequently identified symptoms. A negative net worth was less commonly encountered, with only 11% of the failed sample satisfying this condition.

Reason for Failure	Failed Companies	Non-Failed Companies
(1) No ordinary dividends	88%	41%
(2) No net profit	50%	15%
(3) Negative net worth	11%	1%
(4) Negative working capital	26%	11%
(5) 60% fall in market value	60%	27%

Table 14.3. The percentage of companies in the failed and non-failed samples that met each of the failure criteria. (For complete descriptions of criteria, see previous chapter)

The proportion of the non-failed sample that met each of the criteria is presented in the last column of Table 14.3. A company that meets a failure criterion has not necessarily failed. A failed company, as defined in Chapter Thirteen, also needs to cease existing in its original operational or marketable form (i.e. undergo one of corporate actions listed as A to D in that chapter). The non-failed companies selected in this study had not been delisted, taken over or recapitalised and, so, were not deemed to have failed. Through an assessment of which failure criteria these non-failed companies met, it was possible to gain an understanding of the relative financial health of the non-failed sample in relation to the failed sample.

In terms of the non-failed companies, a nil ordinary dividend was, once again, the most frequently encountered criterion. The remaining criteria are ranked in terms of frequency of occurrence in the same order as that order identified for the failed sample. As expected, the proportion of non-failed companies that met each criterion is significantly lower than the corresponding failed sample proportion.

The graph below illustrates what proportion of the failed and non-failed samples met *x* number of the failure criteria simultaneously. For example, 44% of non-failed firms met no failure criteria, while 4% of the failed firms met all five criteria. The proportion of failed companies that met *x* number of criteria simultaneously rises from nil to two criteria and then declines as the number of criteria increases further. The proportion of the non-failed sample that met simultaneous criteria declines steadily as the number of criteria increases (with no non-failed companies having met all five criteria simultaneously and nearly half the non-failed sample having met no criteria at all).

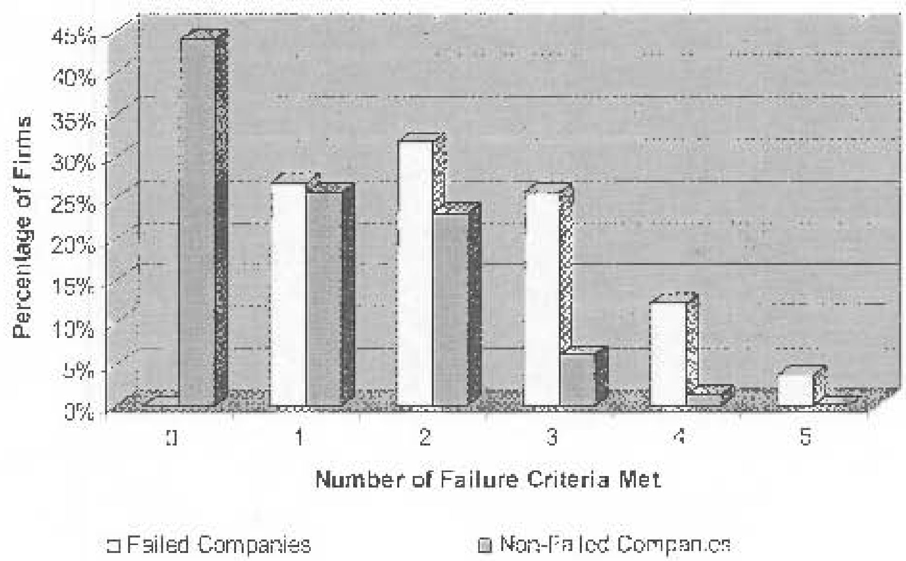


Figure 14.2. Percentage of selected companies simultaneously meeting *x* number of failure criteria.

A **second** sub-sample of failed companies (and their respective non-failed counterparts) was generated so as to represent the **more severe** failure cases. This was performed by eliminating all failed company pairs that met only a single failure criterion. The use of both the full and "severe" samples is described in the following chapters.

CHAPTER 15
FEATURE SELECTION

As stated in Chapter One, the premise of this report is that a series of poor financial decisions lead to the deterioration in the financial health of a firm and finally to its demise. Although the decisions are not directly observable, their consequent effect on the financial health of the firm can be observed.

Financial health can be measured in various ways. However, it should be reiterated that the research objective of this study explicitly states that the model to be constructed should utilise information generally available to investors (see Chapter One).

This study has mainly employed information available from company financial reports. This information is publicly available to investors (only listed companies fall within the scope of this study) and should provide a fair measure for the financial health of each firm.

This chapter begins with a discussion of financial statements and ratio analysis. The full set of features selected for this report is then presented. Finally, the feature set is justified and the calculations of the various variables are explained.

15.1. FINANCIAL STATEMENTS AND RATIO ANALYSIS

15.1.1. FINANCIAL STATEMENTS AND ACCOUNTING INFORMATION

A business is an organisation of people engaged in operations for the purposes of producing economic output.

> "In a capitalistic society, where economic advancement can best be achieved through a strategic analysis of the environment, it is important that the data necessary for rational decision making be made available to those who need it. An essential element in this process is that the financial transactions of the economic unit are clearly documented and presented at specific intervals ... [this is] known as financial reporting, [and] is concerned with the quantitative expression of economic transactions." (Court, 1992, 11)

The financial statements of a company report on this financial performance and position. Financial statements are prepared and presented for use by many different stakeholders throughout the world. In South Africa, the Companies' Act No.61 of 1973 requires that "annual financial statements of a company shall, in conformity with generally accepted accounting practice, fairly present the state of affairs of the company ... and the profit or loss of the company for that year ..." (s. 286(3))

The South African Institute of Chartered Accountants (SAICA) has published a "Framework for the Preparation and Presentation of Financial Statements", AC000. This document forms the underlying basis for all other statements of GAAP. AC000 defines the objective of financial statements as follows:

> "To provide information about the financial position, performance and changes in financial position of an enterprise that is useful to a wide range of users in making economic decisions." (paragraph 12)

AC000 sets out four qualitative characteristics that need to be adhered to in order to make the information provided by financial statements useful to users. These include:

- Understandability
- Relevance
- Reliability
- Comparability

It is an inherent assumption of this study that the financial statement information utilised has been prepared in conformance with the underlying principles of these qualitative characteristics. For example, if the financial statements do not have the attributes of faithful representation, substance over form, neutrality, prudence and completeness, that are associated with the "reliability" characteristic, then much of the data used to construct the failure prediction model in this study will be flawed. As a result, the model itself will be flawed.

Correia, Flynn, Uliana & Wormald (2000, 142) identified further limitations in the use of accounting data:

- **Monetary expression**:
  Information that cannot be expressed in monetary units will not be presented in the financial statements.

This study has included variables for financial market data and the age of the company. In addition, qualitative variables, such as audit report qualification and group holding structure, that are available from the financial statements, have also been included.

- **Simplification and summarisation**:
  Highly complex and diverse economic events need to be recorded in the financial statements. This often requires that such events be simplified and summarised into a presentable format, with the result that vital information may be hidden.

  As a result, the financial data used in this study was thoroughly scrutinized in order to identify any critical information evident in the movement of financial numbers over the years under investigation.

- **Estimation, judgement and accounting policies**:
  A company may make many subjective choices in the selection of accounting policies and in presenting financial information. However, with the increasing acceptance of international accounting standards in South Africa, this level of subjectivity is being reduced.

  In the preparation of data for use in this study, the uniformity in key policies (for example, inventory valuation method) was assessed and accordingly adjusted. In addition, adjustments were made to the data for items that involved a great level of estimation and judgement. For example, over the period under study, accounting practice allowed goodwill to be written off (or not to be written off at all) in a number of different ways. As a result, all goodwill impacts on the balance sheet and income statement were eliminated for the purposes of ratio calculation. A complete discussion of these adjustments is included in the last subsection of this chapter.

- **Inflation**:
  Most accounting information is presented at historic cost. This means that, for example, a fixed asset will appear on the balance sheet at the value at which it was originally purchased less any subsequent amortisation. As a result, in an inflationary environment, this skews conclusions drawn when comparing accounting data across years and companies.

  The data set used in this study include stock market value information in order to build a measure of the true economic value of the firm into the failure prediction model.

Despite these limitations, financial statements are still the most commonly utilised source of financial information by investors. This was one of the reasons that it was considered to be the best source for this corporate failure study.

15.1.2. A "THEORY" OF RATIO ANALYSIS

As Correia et al (2000, 147) noted, the usefulness of a ratio is wholly dependent upon the skilful application of financial data in calculating the ratio and then in the educated interpretation of that ratio.

A common theory underlying the selection of ratios for use in corporate failure studies is the "cash-flow concept" as described by Beaver (1966) in his seminal work. In terms of Beaver's framework, the firm is viewed as a reservoir of liquid assets which is supplied by inflows and drained by outflows. The reservoir serves as a buffer against variations in these flows. The probability that a firm might fail can be defined in terms of the probability that the reservoir will run dry.

This concept can be illustrated as follows:

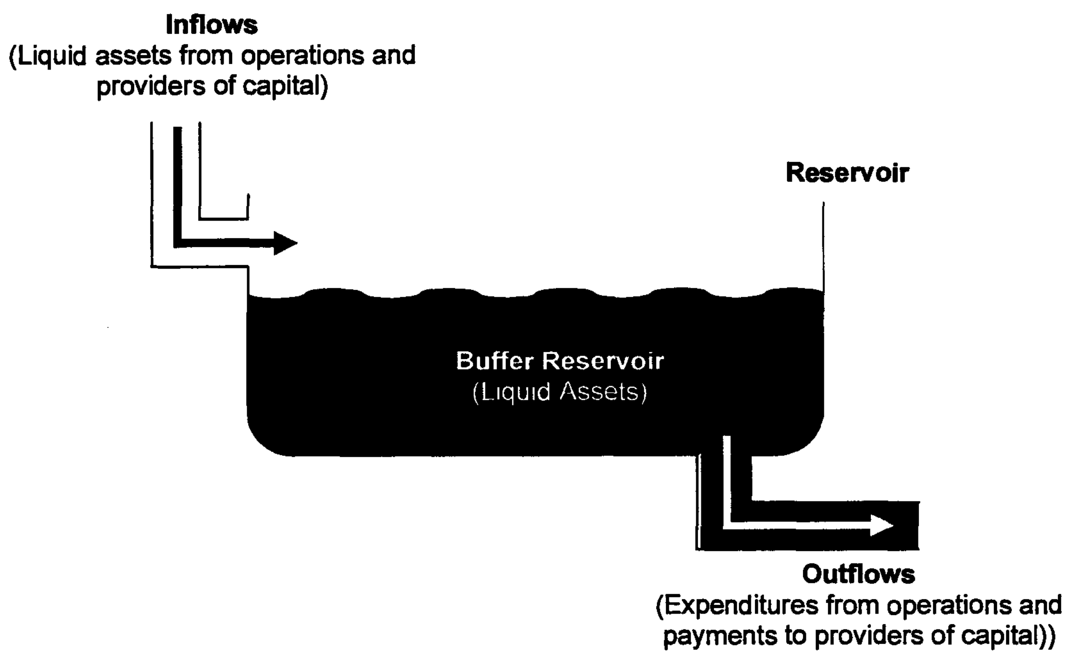


Figure 15.1. The "Cash-Flow" concept illustrated diagrammatically

The relationship between providers of capital and this "reservoir" differs in the cases of equity and debt investors. Both influence inflows (share capital and loans, respectively) and outflows (dividends and interest, respectively), however, the outflow related to debt has a more severe impact on the stability of the reservoir level. This is because in times of financial distress a company can cease dividend payments, while the company remains contractually bound to meet interest payments.

In order to use this model as a vehicle for explaining the ratios selected to predict corporate failure, four concepts need to be highlighted:

- **Size of reservoir**: the larger the reservoir, the smaller the probability of failure.
- **Size of liquid-asset inflow**: the larger the inflows, the smaller the probability of failure.
- **Size of debt**: the greater the debt, the greater the probability of failure.
- **Size of fund expenditures from operations**: greater expenses result in a greater probability of failure.

Additionally, the variability of earnings and claims against resources will influence how "failure-prone" the company may be.

Beaver was of the opinion that any financial ratio for corporate failure prediction can be analysed in terms of its influence on each of the above four impacts on the probability of failure. The ratios selected in this study have been justified on this basis (see below). Table 15.2., below, summarises this selection in terms of the "cash-flow" concept.

Both Beaver (1966) and Blum (1974) used this concept for the selection of predictor variables. Table 5.1. in Chapter Five indicates how Blum combined the cash-flow concept with his three failure-related factors to select the independent variables for his study.

15.1.3. Limitations of Financial Ratio Analysis

Ratios are an excellent tool for assessing and understanding the financial situation of a company. They allow for the better comparison of companies that have different absolute values for financial balances. However, financial ratios suffer from a number of shortcomings that need to be considered.

- **Limitations associated with accounting data**:
  As financial ratios are calculated using accounting data, they suffer from the same shortcomings of financial statements. These have been discussed above and include limitations associated with monetary expression, simplification and summarisation, estimation and choice of accounting policy, inflation, as well as

issues relating to the ethical "fair presentation" disclosure by management of the financial position and performance of the company.

- **Ratios are inter-related**:
  There are numerous inter-relationships between the items in a company's financial statements. As Court (1993, 24) noted, this means that certain sets of ratios may be closely related and a judgement based on composite ratios must be made with caution. He uses the example of high inventory turnover. High inventory turnover may indicate adequate management of working capital or, alternatively, a shortage of goods for sale resulting in the possibility of stock-outs.

  For this reason, a wide range of ratios that span different financial and operational areas, were calculated in this study. An unbiased feature extraction technique, population based incremental learning (PBIL), was then used to find the sets of variables that resulted in the best classification of firms as failed and non-failed. Through such feature extraction, the best combination of inter-related ratios could be selected. This is discussed in detail in Chapters Seventeen and Eighteen.

- **Percentage changes in ratios**:
  If one is to compare percentage changes in ratios, it is important to take into account the absolute values of the amounts underlying the ratios. A 10% increase in asset turnover is significantly more difficult for a firm with a turnover of R1 billion than for a firm with a turnover of R100,000.

  The rate of change of ratios was not included as a potential variable in this study. Rather, all variables were presented to the feature selection technique in ratio format. PBIL then selected single ratios or, alternatively, ratios over a number of years, depending on which most aided classification accuracy. In this way, movements in ratios were accounted for where necessary.

- **Benchmark for comparison**:
  Erroneous conclusions may be drawn when comparing ratios across industries. Argenti (1976) suggested that specialised equations be developed for each section of an industry. However, Immelman (1980) noted that the South African situation largely precludes the implementation of this suggestion because of the limited sample of companies within each sector listed on the JSE.

  Financial services and mining companies were excluded from the sample used in this study. These industries are structurally and operationally the most different to manufacturing/retail/service companies. However, there was still a similar, though less severe, limitation in this study as companies were compared across the industries included in the sample. In addition, the limited sample size of 164 companies did not help to mitigate this problem.

The use of PBIL, the employed feature extraction technique, is a mitigating factor. Through the use of PBIL, the ratios that best discriminate between failed and non-failed can be selected. The features selected should reflect the set of ratios that maximise classification accuracy across the selected sample. In this way, any industry specific ratios, while potentially vital in corporate failure prediction in that industry but not useful in other industries, will be omitted. The best ratios for generalisation across industries should remain.

15.2. SELECTION OF POTENTIAL FEATURES

A number of ratios were calculated using the data collected. The ratios were selected as potential features for training the model to distinguish between failure and non-failure. The process of selecting a relevant subset of features from this set of "potential ratios" is discussed in Chapters Seventeen and Eighteen. The full set of selected "potential" features is discussed in this chapter.

Variables were selected based on a combination of the following considerations:

- Popularity of the ratio in other corporate failure studies (see Appendix B.2.)
- Justification of the ratio in terms of the "cash-flow" concept, as described above.
- Avoidance of the pitfalls of using financial accounting and ratio information in analysing a company's financial health. This has been discussed at the start of this chapter.
- Categorisation of ratios in *Financial Management 4^{th} Ed.* (Correia et al, 2000, 135-165)
- Inspection of the data available from the standardised financial statements on BFA RAID Station (see Appendix C)
- Prior knowledge and evaluation of possible relationships between financial data and failure.

The selected variables are illustrated in the table below. These variables have been grouped into the following categories: liquidity, operating efficiency, operating profitability, solvency, cash flow, market, risk analysis, size and growth, and non-financial data. The variables selected are coded for ease of reference. The coding appears to the left of the variable name. These codes reappear in the appendices and will be referred to later in this study.

The calculation for each ratio is provided in Appendix E.1. The explanations for the abbreviations for the components to each calculation are provided in Appendix E.2. Each explanation makes reference to the source line item number in the standardised financial statements template from BFA RAID Station (see Appendix C).

Each category of ratios selected is now discussed further in order to highlight the rationale behind their selection as potential features for the failure prediction model that has been constructed.

Potential Variables Selected			
	Liquidity	C5	Cash flow to int-bearing debt
L1	Current ratio	C6	Cash flow to total debt ratio
L2	Quick ratio	C7	Cash flow from ops to debt
L3	Cash ratio	C8	Change in LT debt to LT debt
L4	Non-group current ratio	C9	Change in total debt to debt
L5	Non-group quick ratio	C10	Proceeds on shares to assets
L6	Cash to total assets	C11	Cash invested in investing
L7	Receivables turnover		activities to fixed assets
L8	Inventory turnover	C12	Non-cash portion of dividend
L9	Payables turnover		
L10	Cash conversion cycle		**Financial Market**
L11	Preference dividend declared?	M1	Trading Turnover
		M2	Total dividend yield
	Operating Efficiency	M3	Earnings yield
E1	Total asset turnover	M4	Market value of equity to debt
E2	Fixed asset turnover	M5	Price book ratio
E3	Equity turnover	M6	Capital to EBITDA
		M7	Excess return (1 year)
	Operating Profitability	M8	Excess return (2 years)
P1	Operating profit margin		
P2	Net profit margin		**Risk Analysis**
P3	Return on total capital	R1	Business risk (over 3 yrs)
P4	Return on total equity	R2	Sales variability (over 3 yrs)
P5	Return on ordinary equity	R3	Operating leverage (over 3 yrs)
P6	Basic EPS	R4	Business risk (over 2 yrs)
P7	Headline EPS	R5	Sales variability (over 2 yrs)
		R6	Operating leverage (over 2 yrs)
	Solvency		
S1	Debt-equity ratio		**Size and Growth**
S2	Long-term (LT) debt to assets	G1	Log (Assets)
S3	Int-bearing debt to assets	G2	Log (Market Capitalisation)
S4	External LT debt to assets	G3	Retained income to assets
S5	Financial leverage	G4	Retention ratio
S6	Total debt to assets	G5	Growth rate
S7	Debt & contingencies to assets	G6	Capital commitments to fixed
S8	Total commitments to assets		Assets
S9	External debt to assets		
			Non-financial
	Cash Flows	N1	Directors' shareholding
C1	Interest coverage ratio	N2	Auditor's report qualification
C2	Fixed charge coverage ratio	N3	Controlled by another co?
C3	Interest expense to cash flow	N4	Number of years listed
C4	Cash flow to LT debt	N5	Number of years in existence

Table 15.1. All variables selected and calculated as potential features for failure prediction (including their reference code)

15.2.1. Liquidity Variables

Liquidity ratios, in particular the current ratio, are some of the most commonly used variables in corporate health analysis (see Appendix B.2.). Beaver (1966, 71) noted that ratio analysis began at the turn of the 20th century with the development of the current ratio for the purpose of evaluating credit-worthiness.

Liquidity variables focus on working capital balances (*L1* to *L5*). These balances are critical in managing the inflows of liquid assets against the outflows of liquid assets, as described in the "cash-flow" concept above. The inclusion of the cash conversion cycle (*L10*), and its component parts (*L7* to *L9*), is key to assessing the management of the timing of these flows.

Altman (1968, 594) pointed out that a firm experiencing consistent operating losses will have shrinking current assets in relation to total assets (*L6*).

Working capital ratios were calculated including and excluding the more illiquid component (inventory) and the less critical components (inter-group short-term balances). In addition, the cash balance, a vital component of the "buffer reservoir" referred to in Figure 15.1., was also assessed.

The non-payment of preference dividends (*L11*) is considered to be a key indicator of financial distress in many studies. In fact, De la Rey (1981, 11) used this as a critical criterion in the selection of his sample of failed companies. A dummy variable, which relates to the payment of preference dividends, was included in the set of potential variables under investigation. The coding of this label is –1 and +1 for the non-payment and payment of preference dividends, respectively.

15.2.2. Operating Efficiency

This group of variables assesses the efficiency with which assets (*E1* and *E2*) and capital (*E3*) are utilised to generate revenue (i.e. the amount of turnover generated per Rand invested in assets and equity). In this way they measure the efficiency of the firm in maximising the inflows to the cash-flow reservoir.

Altman (1968) included the total asset turnover ratio (*E1*) as one of five distinguishing ratios in his Z-score analysis model.

15.2.3. Operating Profitability

Operating profitability variables assess the ability of the business to generate positive net flows to the reservoir described in Figure 15.1.

These ratios consider the margins on sales (*P1* and *P2*) and the profit returns on various sources of capital. These sources include equity (with (*P4*) and without (*P5*) preference share capital) and total capital (equity plus debt, *P3*).

Basic and diluted earnings per share (*P6* and *P7*), adjusted for share splits in order to maintain comparability, were included as absolute variables. The rationale is that movements in these absolute values across time may indicate a decreasing net flow to the reservoir, while disregarding how much capital is invested in generating that net profit.

15.2.4. Solvency

> "Solvency ratios indicate how management has financed the capital commitments of the firm and accordingly the firm's ability to meet long-term obligations." (Court, 1993, 22)

The various obligations assessed in the set of variables selected include:

- **Total debt** (*S1*, *S5* and *S6*):
  This is important in measuring all probable and measurable obligations to which the enterprise has committed itself.

- **Long-term debt** (*S2*):
  While short-term debt can be rolled over or financed with working capital balances, the repayment of long-term debt is an obligation critical to the continuing existence of an entity in the longer term.

- **Interest-bearing debt** (*S3*):
  This debt is important as it generates additional interest payment commitments in the period leading up to repayment.

- **Debt balances excluding inter-company balances** (*S4* and *S9*):
  It can be argued that the repayment of inter-company debt (as opposed to external debt) is less critical to the continuation of the business as a going concern. This debt can often be subordinated, rolled over or restructured in order to ensure the continued existence of the company.

  The financial statement data obtained from BFA RAID Station was consolidated. Therefore, the inter-company debt balances that were adjusted for, relate to debt held between companies in the group that had not been consolidated (for example associate companies).

- **Debt plus contingencies** (*S7*):
  Contingent liabilities are possible obligations of the firm. The risk to the firm, should these obligations become payable, was assessed.

- **Total commitments of the firm** (*S8*):
  Other than loans and contingencies, the firm may also be committed to capital expenditure or lease payments over the long-term.

  The level of such commitments by the firm may be indicative of failure in two contrary manners. Firstly, if there are insufficient resources to cover the commitments, the firm may be over-extending itself and cash flow issues may arise. Alternatively, a firm that is about to fail will not make capital expenditure and lease commitments for the future, resulting in a lower level of overall commitments. Although these arguments contradict each other, in combination with other variables, the correct information content of these balances will effloresce.

Solvency addresses the impact of the size of debt on the probability of failure as per the "cash-flow" theory described above.

15.2.5. Cash Flows

As the name of the theory suggests, the "cash-flow" concept focuses on the movement of cash. While accrual accounting is important in assessing the performance of a company over a specific period, cash flows are a direct measurement of the actual flows into and out of the reservoir.

The cash flow variables calculated address the coverage of fixed charges and the cash flows to and from different sources.

- **Coverage ratios**: The fixed charges that a firm may face on a periodic basis include interest expense (*C1*), lease payments and preference dividends (*C2*). The selected variables measure the extent to which these charges are covered by cash flows.

- **Cash flows from various sources**: The following cash flows were included in calculated ratios:

- **Cash generated by operations (CF)**:
  This is a measure of the net cash flow to the business from operating activities (i.e. excluding dividend, interest and tax payments). CF was measured against interest expense (*C3*) and debt (*C4* to *C6*) in order to measure the sustainable operating cash flows available for repayment of and servicing of the debt of the company.

- **Cash flows from operating activities (CFO)**:
  This is cash generated by operations (CF) after the deduction of tax, interest and dividends. CFO positive flows are available to the firm for investments in assets or for payments to sources of finance. Conversely, outflows will need to be funded through disinvestments or financed from capital sources.

  CFO was measured against the debt of a company to measure the extent to which this debt could be repaid (*C7*).

- **Cash invested in investing activities (CFIA)**:
  All cash invested in operating assets or return-generating investments was measured against fixed assets (*C11*).

  Companies that are financially sound may invest large amounts of cash in fixed assets in order to rapidly grow the operating asset base. Alternatively, rapid expansion coupled with poor cash flow or working capital management, or even poor growth prospects, could result in the demise of the firm. Once again, the correct information content of this ratio can be better assessed in combination with other variables.

- **Cash flows from debt**:
  The cash flows from the issue or repayment of debt are critical in assessing an expansion or reduction in debt financing (*C8* and *C9*).

- **Cash flows from share issues**:
  A company that issues a significant number of shares may generate sufficient cash to allow for investments in assets, repayment of debt or for investment in working capital. As a result, the financial position of a company may change after a large share issue. This variable (*C10*) seeks to provide information relating to such possible changes in other variables.

15.2.6. Financial Market Data

Financial market variables include information relating to both market value and market sentiment.

Market value information was deemed to be important as most variables employed were based on historical cost. Market value variables added a measure of real value to the feature set describing each company. The market value of the equity of the company, as measured by its market capitalisation, was measured against the book value of that equity (*M5*) and the value of the company debt (*M4*) (an additional solvency measure). In addition, the return on this market value was measured in relation to dividends (*M2*) and various forms of income (*M3* and *M6*).

The sentiment of the market was measured through movements in this market value (*M7* and *M8*). The number of shares traded (*M1*) contributed a measure of the degree of investor interest in the stock.

15.2.7. Risk Analysis

Three risk variables were selected to be included as potential features.

Business risk (*R1* and *R4*) measures the standard deviation of operating income. Variability in income will affect the probability of failure and is a key characteristic of the cash-flow model. The rationale behind sales variability (*R2* and *R5*) is the same as that for business risk.

Finally, operating leverage (*R3* and *R6*) measures the extent to which an increase in sales will result in an increase in operating profit. This is effectively a measure of the impact of fixed costs on the change in profit. Such variability is vital in measuring the probability that the reservoir (of the cash-flow concept) runs dry.

Each measure of risk was calculated over a two and three year period.

15.2.8. Size and Growth

While size was controlled for in the sample selection procedure described in Chapter Fourteen, it was noted that due to certain inherent limitations, companies could not be matched exactly by size.

As a result, three variables measuring the size of the company were included in the set of potential features. The natural logarithm of total assets (excluding intangible assets) (*G1*) and market capitalisation (*G2*) measure the company's accounting and market value size, respectively. The logarithm is taken in order to logarithmically transform the size measures - in this way, the distinction in size between two companies decreases as the absolute size of both increases. This is consistent with the earlier discussion in which it was noted that smaller firms experience an increased probability of failure (Ohlson, 1980, 118).

Altman (1968, 595) noted that the size of retained income will provide an indication of the age of a company. Companies that have been operating for longer periods will often have larger retained income balances that have accumulated over those periods. In addition, the size of retained income reflects the impact that incurred losses will have on this balance. The size of retained income was measured relative to the total assets of the company (*G3*). This is the same ratio that was included in Altman's Z-score model.

The growth prospects of the company were measured using the retention ratio (*G4*) and return on equity (*G5*). These variables measure the return that can be earned on profits that are reinvested in growing the business. In addition, the proportion of capital commitments to fixed assets for each company (*G6*) was included so as to measure the company's own perceptions and intentions regarding its growth potential.

15.2.9. Non-Financial Data

Non-financial data has been shown to be useful in conveying qualitative information regarding the health of a company. Peel et al (1986) summarised such findings, while Merks (1986) investigated the predictive ability of non-financial ratios from a South African perspective. Their findings are discussed in Chapter Five. The following non-financial variables are included in this study:

- **Directors' shareholding** (*N1*):
  Directors are required to disclose their shareholdings in their company. As the directors will have better knowledge than the public of their company's affairs, movements in these holdings may communicate information regarding the

directors' views regarding the prospects of the company. Direct and indirect shareholdings were added together to calculate this variable. Indirect holdings were included because directors probably have influence over these holdings as a result of their insider knowledge.

- **Auditors report qualification** (*N2*):
  Unfortunately, the reasons for audit report qualifications were not available from any of the sources used in this study. However, the fact that an audit report is qualified contains potentially valuable information. Such qualifications can be as a result of going concern issues (a reason directly related to corporate failure) or poor internal controls (an indicator of poor management) or poor financial reporting.

- **Held by a holding company** (*N3*):
  A company controlled by a holding company may have access to emergency resources should the business falter. For this reason, a variable for companies controlled by other companies was included (coded with a +1 if controlled, alternatively with a nil if not).

- **Age of company** (*N4* and *N5*):
  It has been shown that older companies are less likely to fail (see discussions in Chapter Fourteen)

15.2.10. SUMMARY OF RATIO SELECTION IN TERMS OF THE "CASH-FLOW" CONCEPT

The table below describes which components of the "cash-flow" concept are addressed by the different variable groups. These are matched based on the general information content of each group.

"Cash-Flow" Concept Component	Potential Variable Grouping
Size of reservoir	(L) Liquidity (M) Financial market (G) Size and growth
Size of liquid-asset inflows	(L) Liquidity (E) Operating efficiency (C) Cash flows
Size of debt	(S) Solvency (M) Financial market
Size of expenditures from operations	(L) Liquidity (P) Operating profitability (C) Cash flows
Variability of earnings and claims	(R) Risk analysis

Table 15.2. Summary of "cash-flow" concept components addressed by each variable group.

It is important to remember that, while a few of the possible contradictory interpretations of some of the ratios are highlighted above, these situations are not discussed exhaustively. In order to take account of the varying ways in which a ratio may have impacted on corporate failure assessment, a range of ratios that covered all financial areas were selected. The feature subset selection method, discussed in the following section, was then used to select that subset that best discriminated between failure and non-failure.

15.3. CALCULATION OF THE SELECTED POTENTIAL VARIABLES

The formula for calculating each selected variable is included in Appendix E. The discussion below outlines some of the key considerations in these calculations.

15.3.1. Definitions of inputs into variable calculations

Appendix E.2. lays out the inputs that were used for calculating the variables. This appendix defines each input either with reference to the standardised financial statement template from BFA RAID Station (contained in Appendix C) or through an adjustment to these financial statement numbers. These adjustments are discussed below.

- **Adjustments for intangible assets**:
  Intangible assets are divided into goodwill, patents and trademarks, cost of control and other in the BFA RAID Station standardised financials (items 26 to 29, Appendix C). Generally, the value of intangible assets in the open market can not easily be measured. In addition, all these assets, other than patents and trademarks, can not be traded separately to the company.

  For these reasons, intangible assets (excluding patents and trademarks) were removed from all affected balance sheet line items. These affected line items include total assets, ordinary shareholders' equity and all inputs that were calculated using these amounts.

  In addition, all related income statement line items, including intangibles written off, were adjusted in order to remove the impact of intangibles.

- **Adjustments for preference shares**:
  Preference shares are anomalous in that they bear characteristics of both debt and equity. Preference shares are split into redeemable, irredeemable and convertible categories in the BFA RAID Station standardised financial statements (items 9 to 11, Appendix C).

  Convertible preference shares were included in the amount for total shareholders' equity. Redeemable and irredeemable preference shares were included in the long-term debt balance.

  Convertible preference shares are convertible into equity. As a result, their value and characteristics are similar to those associated with equity. However, redeemable and irredeemable preference shares are considered to be more similar to debt in that they are associated with fixed periodic payments. In addition, redeemable preference shares also require a form of capital repayment.

- **Interest expense calculation**:
  Under certain conditions, interest expense can be capitalised to a qualifying asset. In order to measure the burden of interest expense on the companies under study, capitalised interest expense was added onto the income statement interest charge.

15.3.2. Adjustments for Non-Twelve Month Financial Periods

A company may have a financial period of longer or shorter than twelve months. This will occur when a company changes its financial year end. The Companies' Act of

South Africa requires that such period be no shorter than six months and no longer than eighteen months (s. 285(2)).

Comparing the financial information of companies with different financial period lengths is problematic. However, there are certain ratios that should not be impacted by such differences. Ratios that consistent entirely of balance sheet information should, for example, not be influenced. This is because the balance sheet is a measurement of financial position at a point in time and is not measured over a specified period.

In addition, ratios that consist entirely of income statement information and/or cash flow statement data will also not be affected by period length. As long as both the numerator and denominator have been measured over the same period, the ratio between the two should remain consistently comparable.

Therefore, the only ratios that need adjustment are those that consist partly of balance sheet information (measured at a point in time) and partly of income statement or cash flow information (measured over the financial period).

In order to allow for consistent comparison, the income statement and cash flow components of these ratios were adjusted to reflect measurement over twelve months, while the balance sheet items remain unchanged. The applied formula for adjustment was as follows:

$$ISVI = IS \times \frac{12}{Mths} \qquad (15.1.)$$

Where:

- *ISVI* = Income statement or cash flow variable input
- *IS* = Income statement or cash flow line item
- *Mths* = The number of months in the financial period

"Proceeds on shares issued" was the only cash flow item that was not adjusted in this manner. Share issues will generally be once off events. As a result, proportionally adjusting the measurement in order to equate the amount to a twelve month cash flow is not consistent with the nature of the transaction.

The calculations included in Appendix E.1. include the adjustments for period length.

15.3.3. Avoiding Infinity

There is no mathematical solution for:

$$\frac{x}{0}$$

In order to avoid an undefined solution, certain ratios were inverted so that the component that could be zero was in the denominator rather than the numerator.

For example, the interest coverage ratio commonly used by financial analysts is:

$$\frac{EBIT + InterestIncome}{InterestExpense} \qquad (15.2.)$$

However, in a number of cases, the sampled companies did not pay interest during the period, resulting in an undefined solution to this equation. As a result, this ratio was inverted (the answer then became zero, which is more meaningful).

The information content of the variable is not affected through such an adjustment. Such ratios are marked "Inverse" in Appendix E.

15.3.4. Consistency of Data Used

The consistency of data was maintained through the use of a single standardised source for financial statement data, the standardised financial statements from BFA RAID Station. BFA McGregor summarises the financial statement information for each company into this standardised format in order to ensure consistency in the calculation of certain balances.

In addition, all company data utilised was tested in order to check that the value for inventory had been prepared on the FIFO basis.

15.3.5. Calculation of Excess Returns

The excess returns variables (*M7* and *M8*) measure the returns for each company on the JSE Securities Exchange in relation to the JSE market return. Excess returns are deemed to be better than absolute returns for the assessment of how a company is performing relative to the market.

The JSE FTSE/ALSI was used as the proxy for market return. Although this index was introduced in June 2002, it was recalculated retrospectively back to June 1995 by PeregrineQuant and endorsed by the JSE Securities Exchange. Market returns prior to this date were measured using the old JSE All Share Index (ALSI). The JSE FTSE/ALSI is the preferable measure as this is the index that will be available going forward. This impacts on the applicability of the model. These index returns were obtained from I-Net Bridge.

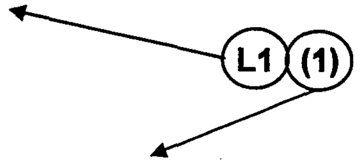
The share price for each company at financial year end was available from the BFA RAID Station database. The return, based on these values, was subtracted from the JSE return. This JSE return was calculated over the actual period that corresponded to the financial period of the company concerned.

15.4. DATA SET CONSTRUCTION AND DATABASE FORMAT

15.4.1. THE 2-YEAR DBASE AND 3-YEAR DBASE

The values of the calculated variables are included in Appendix F. The variables are referenced as follows:

This makes reference to the variable code allocated in Table 15.1.



This refers to the year of the data. A "(1)" refers to variables relating to data from the financial year one year prior to failure.

Figure 15.2. Referencing of variable and year

A number of failed companies included in the sample have only two years of data available prior to failure. As a result, the data set was split into two databases.

The first database (referred to as the "*3-Year Dbase*") included only those failed companies with three years of data prior to failure. This database included the matched three years of data from the respectively paired non-failed companies.

The second database (referred to as the "*2-Year Dbase*") included the two years of data prior to failure for **all** failed companies. Consequently, all non-failed companies were also included in this database. However, only their two years of data, corresponding to the years of their paired failed companies were included.

As a result of the *2-Year Dbase* only including variables for the two years of operation prior to failure, certain features included in the *3-Year Dbase* had to be excluded. Appendix F clearly indicates where this is the case. For example, business risk and sales variability had to be measured over a period of two and three years for the *2-Year Dbase* and *3-Year Dbase*, respectively.

15.4.2. DATA SET CHARACTERISTICS

The following table summarises some of the characteristics of the *2-Year Dbase* and *3-Year Dbase*.

Characteristic	2-Year Dbase	3-Year Dbase
Years of data prior to failure	2 years	3 years
Number of failed companies	82	61
Number of potential features	120	179

Table 15.3. Characteristics of the *2-Year* and *3-Year Dbases*.

CHAPTER 16
DATA PRE-PROCESSING

16.1. DEFINING DATA PRE-PROCESSING

> "Since neural networks can perform essentially arbitrary non-linear functional mappings between sets of variables, a single neural network could, in principle, be used to map the raw input data directly onto the required final output values. In practice, for all but the simplest problems, such an approach will generally give poor results for a number of reasons... For most applications it is necessary first to transform the data into some new representation before training a neural network." (Bishop, 1995, 295)

Bishop went on to note that pre-processing is one of the most significant factors in determining the performance of the final system. He discussed the following forms of data pre-processing:

- **Incorporation of prior knowledge**:
  Information relevant to developing an optimal model should be included in constructing that model. Bishop noted that such prior knowledge can be incorporated into either the network structure (discussed in later chapters) or into the pre-processing of the data (discussed below).

- **Accounting for missing input values**:
  There are a number of potential approaches to dealing with such deficiencies. The approach adopted in this study is discussed below.

- **Linear transformation of the data**:
  This entails linearly rescaling the input data. The data used in this study was normalised as described below.

- **Making adjustments for incorrect target values**:
  If a model is to be accurate in classifying a firm as failed or non-failed, it should at least be able to accurately classify the companies that have clearly failed or not failed. If the model is incapable of classifying these clear cut cases, the model will be less likely to correctly classify the companies that are on the borderline of failure

and non-failure. As such, a data set of the clear cut cases was tested initially. The all-inclusive model was then developed. This is discussed below.

- **Dimensionality reduction**:
  The reduction in data dimensionality was achieved through feature selection (discussed in the following chapter).

16.2. THE INCORPORATION OF PRIOR KNOWLEDGE AND DATA SET CONSTRUCTION

16.2.1. VARIABLE SELECTION

Using prior knowledge - regarding the manner in which the different financial and operating factors of a company impact on the assessment of its financial health - a relevant set of potential variables was selected. The application of this understanding to the calculation and selection of such variables is discussed in the previous chapter.

16.2.2. VARIABLE CALCULATION

Variables presented to the model in ratio format provide an additional form of pre-processing. Ratios eliminate the problems in comparing companies that have different absolute values for financial balances. This is discussed in detail in the preceding chapters.

16.3. DEFICIENT DATA

Deficient data include:

- variables that cannot be accurately calculated as a result of missing financial information, and
- undefined variables as a result of dividing by a zero balance in the calculation of that variable.

These are common problems encountered in empirical research. There are a number of approaches that are proposed for dealing with such missing inputs.

Bishop (1995, 301) noted that where the quantity of data available is sufficiently large, and the proportion of companies affected is small, then the simplest method is to discard those companies with missing data from the data set.

This approach was not possible in this study for the following reasons:

- The proportion of companies impacted by deficient data was 23%. As the sample was already limited in size, the exclusion of such a large proportion of companies was not deemed to be appropriate.
- In certain instances, the fact that data was deficient communicated its information in itself. For example, where a ratio was undefined as a result of dividing through by zero, the exclusion of that variable would modify the data distribution relating to the financial balance that had that zero value.

Therefore, an approach of "filling in" deficient data was adopted. In this way the sample size could be maintained.

The value inserted depended on whether the variable was deficient as a result of missing data or as a result of an undefined solution.

16.3.1. MISSING INFORMATION

Variables that could not be calculated as a result of missing data were replaced with the average value for that variable.

However, replacing a missing variable in such a way impacts on the information contained in the trend of that variable in relation to the value(s) of other variable(s). This is clearly illustrated below.

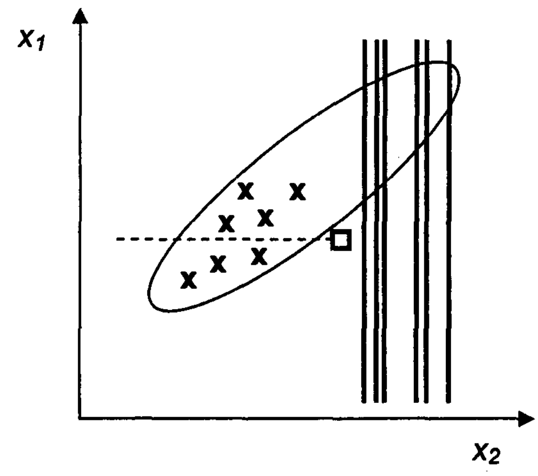


Figure 16.1. Schematic illustration of a set of data input points in two dimensions. For some of the data points (shown by the crosses) the values of both variables are present. For the others (shown by vertical lines), only the values of x_2 are known. If the mean vector of the distribution is estimated using the available values of all variables, then the result is a poor estimate, as indicated by the square. (Source: Bishop, 1995, Figure 8.3.)

In this study, averaging variables across companies and years would have resulted in the loss of such information.

In order to adjust for this problem, the data was segregated into distinguishable portions for the purposes of calculating averages and replacing missing variables.

(a) Failed companies

Average values for affected variables were calculated across failed companies (excluding those where the variable was missing) separately for each year prior to failure. Missing variables were then replaced by the average value of the variable for that specific year prior to failure.

Averaging was performed separately for the sub-sample of failed companies and their non-failed counterparts – this is because of the different information contained within the averages of these two groups. In addition, it was considered necessary to average the data for the failed companies based on **year prior to failure** as the characteristics of the variables in each year should differ as failure approaches.

(b) Non-failed companies

As a non-failed company is not approaching failure, the variable should bear similar information across all the years for which it is included in the model. For this reason, missing data was replaced by the average value for that variable as measured simultaneously across all three years of data.

(c) Illustrative example

For example, missing inventory turnover for a failed company for the year before failure was replaced by the average inventory turnover over all failed companies one year prior to failure. However, if inventory turnover was also missing for the paired non-failed company, that value was replaced by the average inventory turnover as calculated over all inventory turnover values of the non-failed sample, irrespective of year.

The variables that are impacted by missing financial information are listed in the table below. Missing data included information regarding fixed assets and inventory. The cash conversion cycle for companies impacted by missing inventory numbers was recalculated using the replacement value for inventory turnover.

16.3.2. Undefined Variables

When one divides any number by zero, the answer is undefined. As the denominator in a ratio tends toward zero, so the value of that ratio tends toward infinity. Hence, undefined variables were replaced by some maximum value of that variable.

The maximum value was calculated by segregating the sample in the same way as described in 16.3.1. This value was subjectively chosen based on the maximum value of the variable as measured across the relevant portion of the sample.

The following variables used in this study were impacted by this "filling in" process of using average and maximum values.

	Missing variable replaced with average	Undefined variable replaced with maximum
Variable impacted	(L8) Inventory Turnover (E2) Fixed Asset Turnover (C11) Cash invested in investing activities to fixed assets (G6) Capital commitments to fixed assets	(C4) Cash flow to long-term debt (C8) Change in long-term debt to long-term debt

Table 16.1. Variables impacted by missing financial information and zero denominators

16.3.3. Adjustment for Potential Outliers

In other studies of this nature, a subjective manual adjustment has been made in order to cap the outliers of variables. Preliminary investigations using Population-Based Incremental Learning (PBIL) and the Separability Index (SI) indicated that there was no advantage to such a procedure in this study. As a result, in order to minimise the manipulation of data, no outlier adjustment was made.

16.4. NORMALISATION OF VARIABLES

16.4.1. Continuous Variables

The size of the variable inputs should not have a bearing on the relative importance of that variable.

Using a linear transformation, inputs can be scaled to be within a similar range. In this way, no single input would have a greater influence on the model than another.

The data in this study was transformed through the normalisation of the variables. In order to do this, the mean ($\bar{x}_i$) and variance (σ_i^2) for each of the variables (x_i) was independently calculated.

$$\bar{x}_i = \frac{1}{N}\sum_{n=1}^{N} x_i^n \tag{16.1.}$$

$$\sigma_i^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_i^n - \bar{x}_i)^2 \tag{16.2.}$$

Where:

- n = 1, 2, ..., N are the labels for the companies in the sample (both failed and non-failed).

This calculation was performed across all companies (failed and non-failed) and years. In this way, the distinctions between failed and non-failed companies, and between the years prior to failure, were maintained. As noted in the previous chapter, two databases of variables were developed (*2-Year Dbase* and *3-Year Dbase*). This normalisation process was repeated separately on each database. The process is described in Figure 16.3. below.

The rescaled variables ($\tilde{x}_i$) could then be defined by the following equation:

$$\tilde{x}_i^n = \frac{x_i^n - \bar{x}_i}{\sigma_i} \tag{16.3.}$$

After the normalisation process, each variable has a zero mean and a unit standard deviation over the entire sample.

Bishop (1995, 299) noted that, in the case of a radial basis function networks, it is particularly important to normalise the variables in order that they span similar ranges. This is because the basis function is determined by the Euclidean distance, *d*, between the input vector, *x*, and the basis function centre, u_j, given by:

$$d^2 = \left\| x - u_j \right\| = \sum_{t=1}^{d}\left(x_i - u_{ji}\right)^2 \tag{16.4.}$$

where *d* is the dimensionality of the input space. If one of the input variables is scaled, for example, to within a smaller range than the other variables, then the value of d^2 will be smaller and, consequently, the variable will have a different impact on the classification algorithm.

This argument applies equally to any classification technique that is based on Euclidean distance, including the k-Nearest Neighbours and Kernel Ridge Regression algorithms used in this study. Greene (2001) illustrated this problem graphically using a nearest neighbour classifier.

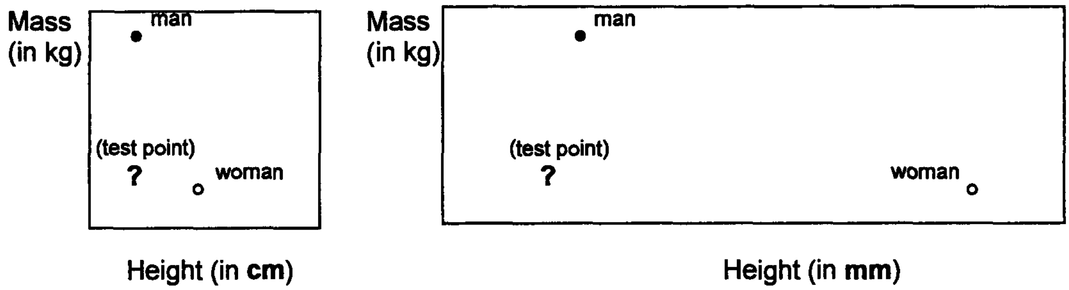


Figure 16.2. Impact of scaling of variables: In the diagram on the left, the closest training example to the test point (?) is a woman; changing the scale units of the height measurement stretches the diagram so that a man is the closest training example.

16.4.2. Discrete Variables

Categorical variables, such as audit report qualifications, were coded in binary. For example, if a report was qualified it was coded with a 1, and if not then it was coded with a 0. As these variables were already within the range 0-1, they did not need to be rescaled.

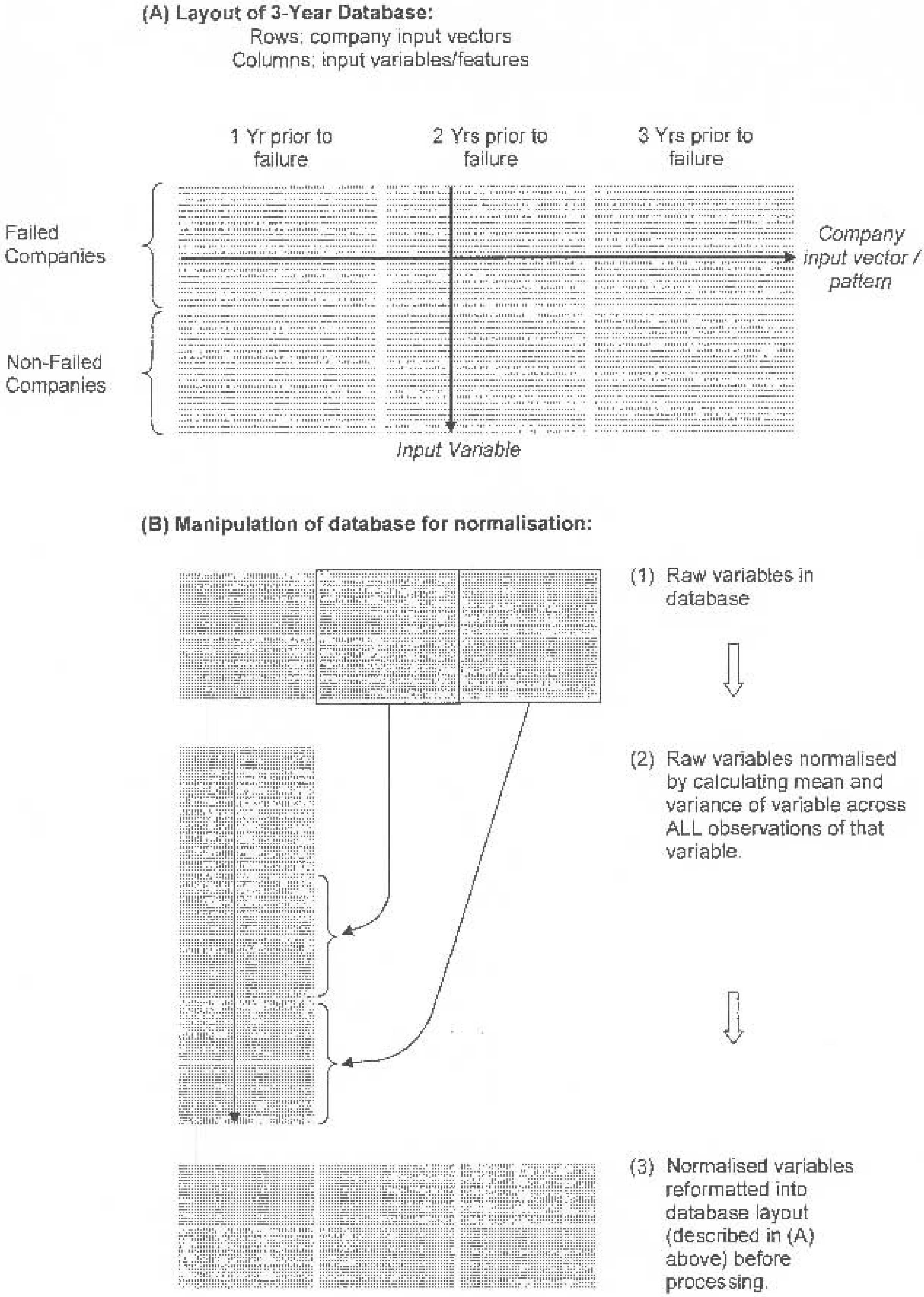


Figure 16.3. Layout of *3-Year Database* and process of manipulating the database in order to normalise the variables.

16.5. ADJUSTING THE DATABASES FOR BORDERLINE CASES (CREATING A REDUCED SAMPLE)

If the final model is to accurately classify those firms that are on the brink of failure or non-failure, then it should be able to classify the clearer cut cases with significant accuracy.

As is noted in Chapter Fourteen, 27% of the selected failed companies met only a single failure criterion. These companies were defined as "borderline failures". In addition, one of the non-failed companies met four out of the five possible criteria for failure (and ultimately did not fail – "borderline non-failure").

In order to develop a database that contained input vectors for only those companies that were clearly failed or non-failed, these aforementioned borderline cases were removed from the full database. The data in this additional reduced database was then tested. The results are discussed in the following chapter.

16.6. SUMMARY OF DATABASES AT THIS POINT

In summary, there are four different databases that were taken into the feature subset selection process. There is a ***2-Yr Dbase*** and a ***3-Yr Dbase*** for the ***full*** and ***reduced*** data sets.

- The ***full*** data set includes all the companies selected as outlined in Chapter Fourteen. These companies are disclosed in Appendix D.
- The ***reduced*** data set is the subset of the full data set that excludes those borderline cases identified as described above. These borderline company pairs can be identified from the list in Appendix D as any pair where either the failed company met a single failure criterion or the non-failed company met at least four.

- The ***2-Yr Dbase*** includes all data for those failed companies (and their non-failed pair) that had two years of data available prior to failure.
- The ***3-Yr Dbase*** includes those companies with three years of available data.

These data items are clearly identified in Appendix F.

SECTION C
EMPIRICAL MODEL CONSTRUCTION: FEATURE SUBSET SELECTION

CHAPTER 17
METHODOLOGY FOR FEATURE SUBSET SELECTION

There has been a growth in the use of machine learning for business and financial applications. However, their use is disadvantaged by the fact that such algorithms can be slow due to the size of the search space involved and the iterative manner in which many of the algorithms function. Piramuthu et al (1998) showed that the degree of difficulty involved in training a neural network is inherent in the set of features presented to the algorithm. They showed that through "feature construction", which involves data pre-processing (discussed in the previous chapter) and dimensionality reduction (feature subset selection), both learning speed and classification accuracy can be improved.

Learning speed does not fall explicitly within the scope of this study's research problem. However, it was deemed to be an important consideration due to the implicit limitations on computing time available for the completion of this research.

Classification accuracy was, however, directly addressed. This chapter first discusses the key aspects of feature subset selection that need to be addressed in order to improve such accuracy. The chapter then progresses on to a justification and explanation of the Population-Based Incremental Learning (PBIL) and Separability Index (SI) methodologies applied in this study.

The following chapter then discusses the application of these techniques to the compiled databases of company information.

17.1. FEATURE SUBSET SELECTION USING BINARY MASKS

Before discussing the approaches to feature subset selection, it is necessary to clarify "binary mask selection".

In this study, a subset of features was selected in the form of a binary mask (with as many elements as there are possible features). The mask was then applied to a vector of features describing a firm in order to determine the selected feature subset. A "1" in the mask included the variable in the subset, while a "0" resulted in the omission of that variable.

For example, a mask of {0, 1, 1} applied to a set of three features {a, b, c} resulted in the subset {b, c}.

17.2. FEATURE SUBSET SELECTION AND GENERALISATION

Generalisation was defined by Zinilli (1997, 8) as the "property ... whereby a [model] is able to provide a correct matching of output data to a set of previously unseen input data." This property is illustrated well in Chapter Nine.

In order to find the optimal balance between a model that is too simple to accurately classify and one that overfits and generalises poorly, the parameters of the model need to be adjusted. Bishop (1995, 332) referred to this as "structural stabilisation". Structural stabilisation entails:

- adjusting the methodology applied using a suitable adjustment method (discussed in the following chapters), and
- selecting a set of features appropriate to better generalisation.

This chapter addresses the latter. Structural stabilisation from the perspective of the applied methodology is discussed in Section D.

17.3. THE IMPORTANCE OF FEATURE SUBSET SELECTION

A potential weakness associated with classification methodologies based on measurements of Euclidean distance, such as the k-Nearest Neighbour (kNN) or Kernel Ridge Regression (KRR) methods used in this study, is their vulnerability to highly correlated features (redundancy) and irrelevant features (irrelevance). Many researchers have made reference to these key issues of redundancy and irrelevance (Cantu-Paz, 2002; Bishop, 1995).

These issues can be addressed through a process that reduces the dimensionality of the data by selecting a subset of the available features, as discussed further below.

17.3.1. MULTICOLLINEARITY

The situation in which the input features are highly correlated with one another is known as multicollinearity. Van den Honert (1997, 139) noted that although the inclusion of highly correlated features may increase the explanatory power of the model, it is preferable, as far as possible, only to include those features that are uncorrelated.

In data modelling, such as in this study, multicollinearity has several pernicious effects, which include:

- Greene (2001b) explained that it reduces the rank of the correlation matrix and results in an unfavourable condition factor. This makes the matrix difficult to invert. Matrix inversion is a key procedure in the application of an inductive technique such as KRR.

- Multicollinearity also indicates feature redundancy, which results in highly correlated features receiving spurious alpha weightings in a regression equation. The extent of multicollinearity in the full feature set used in this study is discussed in the following subsection. This problem, associated with constructing classification models that use highly correlated input data sets, is illustrated graphically in Chapter Twenty in the explanation of ridge regression.

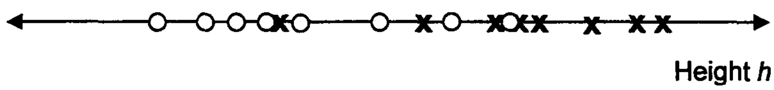
There are a number of methods that have been used in corporate failure studies in order to address the issue of multicollinearity. These include methods such as Principal Component Analysis (PCA) that seek to remap the features into a space defined by a reduced number of orthogonal factors (see Chapter Seven for examples of the application of PCA).

In this study, the effects of multicollinearity were reduced by the feature selection process and further mitigated by the employed regularisation method (ridge regression) (see Chapter Twenty).

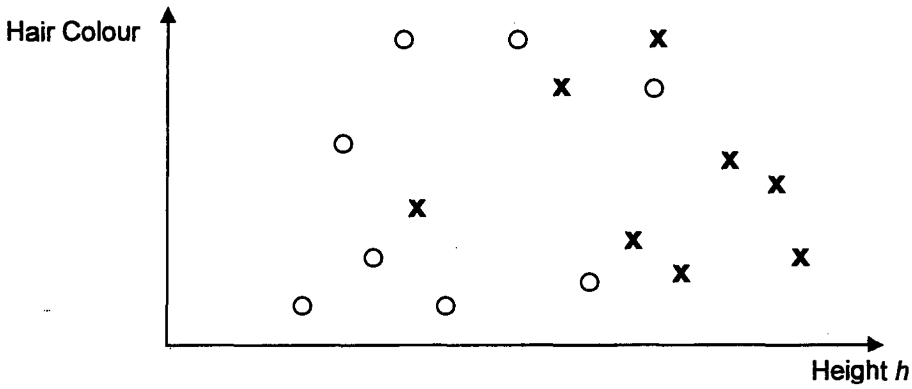
17.3.2. Irrelevant Features

Two classes of data are highly separable in a situation in which, in *n*-dimensional space, such classes of do not overlap. The inclusion of features that do not contribute information to distinguishing between classes will adversely affect the separability of the data. The following example, adapted from Greene (2001b), illustrates this point graphically:

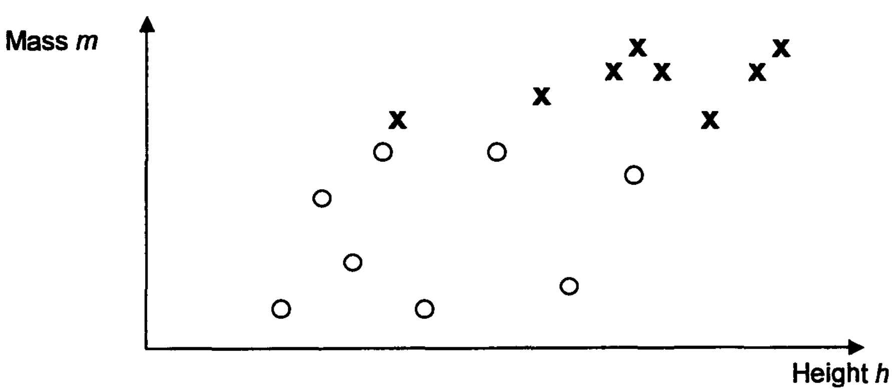
> In trying to algorithmically distinguish between men and women, height may be considered to be a key feature. Plotting men (crosses) and women (circles) on a "height" axis may produce a result that appears as follows:



Attempting to distinguish genders further based on a random feature, such as hair colour, may cause the groups to become inter-mingled and impair the separability of the data.



On the other hand, the inclusion of a gender-relevant feature (for example, mass) could improve the separability of the data and, hence, classification accuracy.



17.4. PRELIMINARY ASSESSMENT OF MULTICOLLINEARITY IN THE FULL FEATURE SET

As is noted in the discussion in Chapter Fifteen, the set of potential features selected for this study was chosen ignoring any information that may have been double-counted in their selection. Rather, it was recognised that, while many features may be highly correlated, there may be unique information contained in each that may provide improved explanatory power when viewed in combination with other features. As a result, multicollinearity is pervasive throughout the full set of selected "potential" features (i.e. before feature subset selection).

17.4.1. "POTENTIAL" VARIABLE CORRELATION MATRIX

A preliminary assessment of the extent of multicollinearity in the potential feature set was performed. A correlation matrix (based on the *2-year Dbase*) was calculated and is included in Appendix G.1. The two-tailed t-test was then used to test these correlations for significance.

$$t_{obs} = \rho_{xy} \cdot \sqrt{\frac{n-2}{1-\rho_{xy}^2}} \qquad (17.1.)$$

The results of these significance tests are included in Appendix G.2.

17.4.2. BRIEF ANALYSIS OF CORRELATION MATRIX

The results of the t-tests indicate that there is significant multicollinearity within the potential feature set.

A brief preliminary analysis of the results was performed in order to get an idea of the information content and relationships between different variables prior to the final subset selection. The following was noted:

- The potential variables were grouped according to information category (see Table 15.1.). As a result, intra-group correlations were expected.

  This was, for example, clearly evident in the correlations between the different forms of the current ratio. The basic current ratio was adjusted for illiquid and inter-group working capital balances to form an additional five variables (*L1* to *L6*). All these variations are significantly correlated with one another at a 1% significance level.

  However, it is interesting to note that within the cash flow, operating efficiency and risk analysis groupings there are little to no significant correlations. The only significant correlations within the cash flow grouping are between the variables that measure cash flows relative to debt (*C6* to *C9*). The lack of intra-group redundancy between these variables can be attributed to the different cash flows, areas of operation and risks areas addressed by the selected variables in each category, respectively.

- Significant correlations across groupings are also evident. In particular, the solvency variables are highly correlated with the current ratio-related variables and certain operating profitability ratios.

  The relationship between solvency and liquidity is a well documented one.

  The significant negative correlations between the solvency and profitability features (measured in terms of margins, returns and earnings per share) imply that as solvency weakens, profitability declines. This may be that the weakening solvency of a firm acts as an indicator of declining financial health. Alternatively, the increased interest and fixed charges associated with larger financial commitments may directly impact the bottom line.

  The investigation of such relationships is outside of the scope of this study. Such analysis is suggested as an area for further research in Chapter 23.

- It is interesting to note that the size features are significantly correlated to various features across all the groupings of variables. This is consistent with findings in prior studies that size is a pervasive factor in corporate failure. This, once again, justifies controlling for size in the selection of the sample (see Chapter Fourteen).

- There are certain groups of variables that have little or no correlation to any other features. These features, with unique information content, include the turnover ratios (liquidity grouping) and coverage ratios (cash flow grouping).

While it was possible to perform an extensive analysis on these correlations in order to draw conclusions about the information content of different features, this was not within the scope of this study. It is suggested as an area for further research (Chapter 23).

It is evident that many variables, with similar information content, were included in the full feature set. It was, therefore, necessary that the subset selection method employed would identify those variables that, alone, held the necessary information content. This was addressed by the methodologies used in this study.

17.5. JUSTIFYING THE PROCEDURES FOR FEATURE SUBSET SELECTION

Any procedure that is used for feature subset selection must be based on two principles (Bishop, 1995, 304):

- A **criterion** must be defined by which it is possible to judge whether one subset of features is better than another. Bishop suggested that this criterion should be similar to that on which the eventually completed model will be assessed. This is a wrapper approach and is discussed below.
- A **systematic procedure** for searching through the various combinations of possible subsets must be found.

17.5.1. CRITERIA FOR FEATURE SUBSET EVALUATION

Exhaustively searching for the best feature subset is futile when there are as many variables to choose from as there are in this study. Given such a situation, there are two types of approaches that can be adopted – the **filter** and **wrapper** approaches (John, et al, 1994).

Cantu-Paz (2002) described the **filter** process as one in which features are selected based on properties that "good features" are presumed to have, such as orthogonality and high information content. This is performed independently of the classifier in which the feature subset is to be used. He noted that although the filter approach can be fast, it may produce disappointing results because it ignores the induction algorithm that will eventually be applied.

A number of different filter-type approaches are discussed in Chapter Seven. In particular, Eisenbeis (1977, 885) and Eisenbeis, Gilbert & Avery (1973) summarised many different measures that can be used in assessing the significance of a variable subset. These include Wilk's lambda and F-ratios.

The key idea of the **wrapper** approach is to use the actual induction algorithm as the method with which to evaluate each feature subset. This can be done through a heuristic search algorithm (John, Kohavi & Phleger, 1994). Although the results obtained using this approach may be superior to the filter approach because the "selection process is attuned to the specific inductive bias of the classifier used" (Greene, 2002), there are two major drawbacks to its implementation:

- it is computationally intensive, and
- the resulting classifier should be tested on the data not used during the search (a problem in situations of limited data availability).

With this in mind, the selection of Thornton's Separability Index (SI) can be justified as a wrapper-type approach that seeks to overcome the aforementioned drawbacks. The Separability Index gives a measure of the separability of data in *N*-dimensional space. This is discussed further in later sections of this chapter.

17.5.2. Systematic Procedures for Subset Searching

The traditional statistical methods employed in searching for an optimal feature subset have been discussed in Chapter Seven. Eisenbeis et al (1973) addressed these in their article entitled "Investigating the relative importance of individual variables and variable subsets in discriminant analysis".

Inza, Merino, Quiroga, Sierra & Girala (1999, 7) discussed the search methods that have arisen with the advent of machine learning. They identified three basic issues, other than setting a criterion for subset evaluation (above), that must be addressed in order to define the nature of the search process:

- defining the starting point,
- how to organise the search, and
- defining the criteria for halting the search.

(a) The starting point

The researcher should select a subset of features with which to begin the search. This initial subset is then adjusted in order to search for the optimal subset.

(b) The organisation of the search

Search strategies can either be complete or heuristic in nature. An exhaustive search is usually unfeasible as there are 2^N-1 possible subsets in a search space of N possible variables. The computing time required for such a search in this study would exceed the longest of human life expectancies!

Greene (2001) identified four types of heuristic searches.

- A **random search** randomly selects different subsets until a termination criterion is satisfied.
- **Greedy searching** sets all bits in the mask to "1" and systematically changes each bit to "0". If there is an improvement in the evaluation criteria, then the change is retained. Otherwise the bit is reset to "1". This approach is similar to stepwise regression and suffers from the same shortcomings (see Chapter Seven).
- **Stochastic hill-climbing** is similar to a greedy search, except that the starting point is a randomly generated mask and random bits are changed in order to see if there is an improvement in the evaluation criteria.
- **Evolutionary searches** take note of the masks that produce the better results. New randomised bitstrings are then generated so as to resemble

these earlier discovered masks. In this way, the search focuses progressively on that area of the search space that maximises the separability index. Genetic algorithms (GA) and population-based incremental learning (PBIL) are examples of such strategies.

(c) *Criteria for halting the search*

If a heuristic search method is used, criteria will need to be set for stopping the search. These may relate to a non-improvement in the evaluation criterion or to simply terminating the search after a set number of iterations.

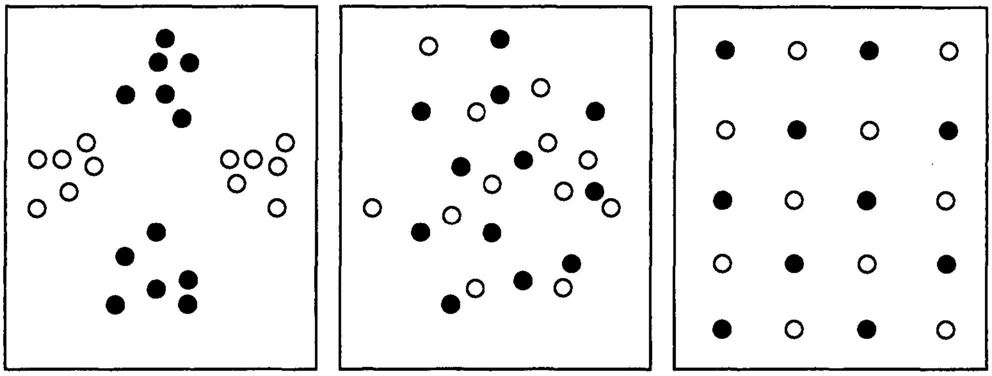
This study employed Population-Based Incremental Learning (PBIL) with a randomised bit-mask starting point. The optimal number of iterations over which to run PBIL was determined through trial and error. PBIL is discussed later in this chapter.

17.6. THORNTON'S SEPARABILITY INDEX AS A CRITERION FOR EVALUATING FEATURE SUBSETS

17.6.1. DISCUSSION OF THORNTON'S SEPARABILITY INDEX

Thornton devised a measure for rapidly assessing the degree to which a set of points can be classified on a geometric basis. He defined the separability index (SI) as equal to the percent training points which share a classification label with their nearest neighbour (Thornton, 1997).

Greene (2001) illustrated this concept with the following example.



SI = 1
Every point has a nearest neighbour of the same class

SI = 0.5
About *half* the points have a nearest neighbour of the same class

SI = 0
No point has a nearest neighbour of the same class

Figure 17.1. Illustration of different separability index values (Source: Greene, 2001b)

The data points in the box on the left comprise well-separated clusters. Therefore, most points will have the same classification as the points around them (except, perhaps, for the points on the edges of the clusters). SI will be close to 100% (SI = 1). The randomly distributed points in the middle box have no clear geometric separation while the points in the box on the right are adjacent to points of an opposing class.

SI does not appear to have been used to any extent in academic research (a search on www.citeseer.com yielded no results other than Thornton's original paper and subsequent book). However, it does appear to have merit as a fast and easy way to judge the classification accuracy of a subset of features to be used in a sparse proximity-based classifier (such as kNN and KRR).

17.6.2. Justification for the Use of Thornton's Separability Index

As noted above, the wrapper approach to feature subset selection generally yields better results than the filter approach. Greene (2001a), in his paper entitled "Feature subset selection using Thornton's separability index and its applicability to a number of sparse proximity-based classifiers", justified the SI by comparing it to a wrapper-based approach that used a nearest neighbour classifier. He noted that, in order to maintain generalisation accuracy using such a nearest neighbour approach, the computational burden of the train/test requirements are great.

However, he found that the SI "yields a result identical to the asymptotic result of a large number of train/test cycles with random splits, but it requires only a simple non-iterative calculation" (Greene, 2001a, 2).

In addition, Greene also found that the SI is effective in evaluating feature subsets to be used in kernel-weighted induction algorithms.

The SI, thus, enables the evaluation of feature subsets using a method that is based on proximity, without the computational drawbacks of the standard wrapper approaches. The induction algorithms used in this study (kNN and KRR) are also sparse proximity-based algorithms, making the SI a useful substitute for the superior wrapper approach.

17.6.3. CALCULATION OF THE SEPARABILITY INDEX

The Matlab code that was used to run the SI is included in Appendix A.1. A description of the logic behind the code is included in this appendix.

17.7. POPULATION-BASED INCREMENTAL LEARNING (PBIL) AS AN APPROACH TO FEATURE SUBSET SELECTION

17.7.1. BACKGROUND TO DEVELOPMENT OF PBIL

Population-Based Incremental Learning (PBIL) was first formulated by Baluja in 1994. However, it has its roots in genetic algorithms. In fact, a seminal paper published on PBIL by Baluja and Caruana in 1995 was entitled "Removing the Genetics from the Standard Genetic Algorithm".

> "Genetic algorithms are one of the best known techniques for solving optimisation problems." (Inza et al, 1999, 9)

The genetic algorithm is a population-based search method. Initially, a genetic algorithm generates a random population of binary vectors. Each vector is evaluated and the better performing bitstrings identified. A new population of binary vectors is then generated so as to bear some resemblance to these "performers". These subsequent vectors are generated through processes including selection, recombination and mutation (Cantu-Paz, 2002). Through these processes, the genetic algorithm can implicitly extract, from each generated population, statistics about the better areas of the search space (Baluja, 1996, 1).

Genetic algorithms (GAs) have been used in many studies to obtain "structural stabilisation" (discussed at the start of this chapter). GAs have been used in both optimal feature subset selection (Cantu-Paz, 2002; Inza et al, 1999; Back, Oosterom, Sere & Van Wezel, 1995 and Back, Laitinen & Sere, 1996, just to name a few papers that were reviewed in completing this study) and in selecting parameters for neural network architecture choices (Hansen, 1998 and Welch et al, 1998).

PBIL is a combination of GAs and competitive learning (Baluja, 1994). The key difference between PBIL and GAs is that PBIL attempts to *explicitly* maintain statistics about the better areas of the search space through a probability vector (GAs maintain this information *implicitly* through mutation and recombination of the better performing binary vectors) (Baluja, 1996).

17.7.2. Description of the Basic PBIL Algorithm

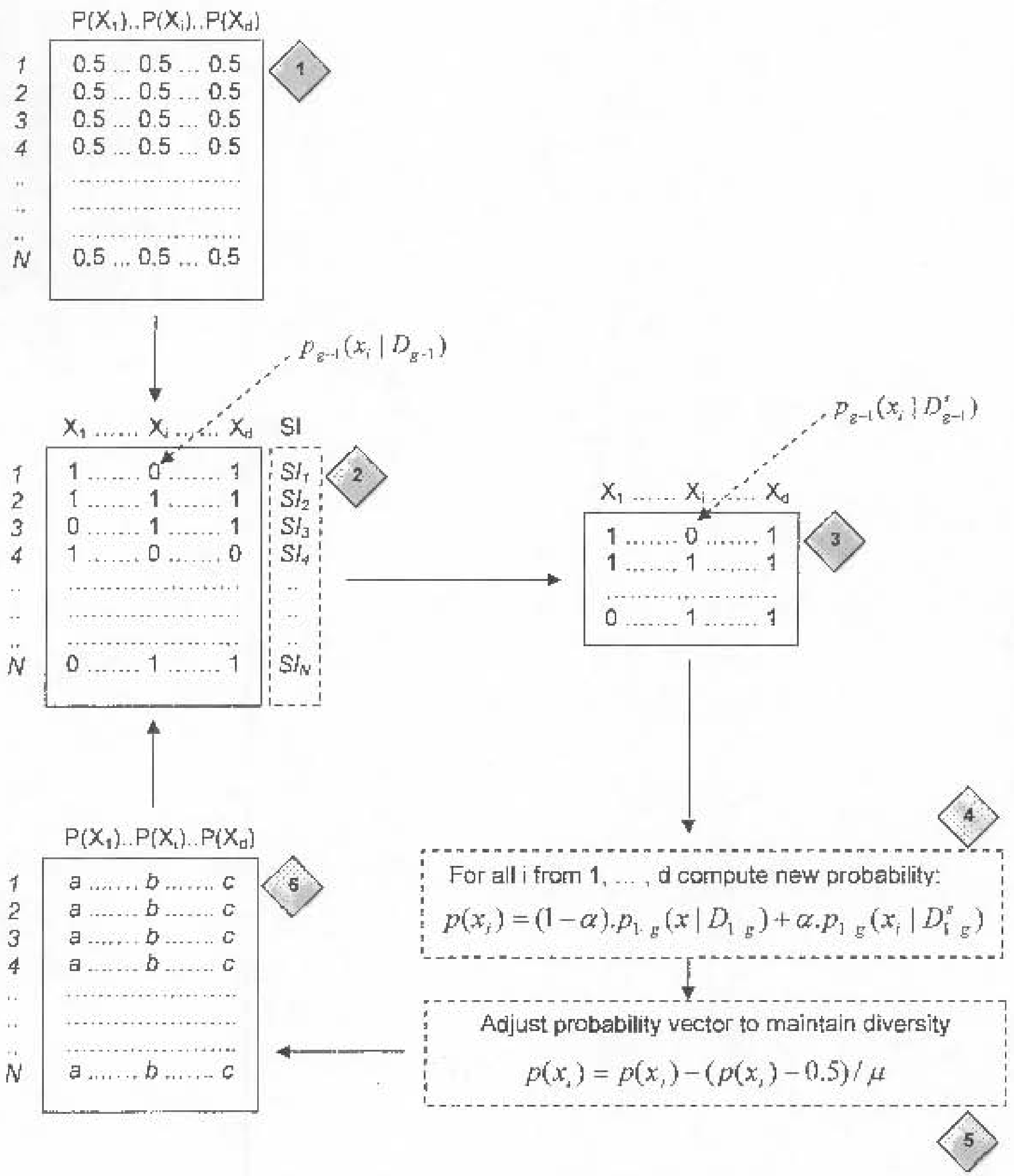


Figure 17.2. Diagram of the basic PBIL algorithm used in this study

The above figure was adapted from Inza et al (1999, 11). It illustrates the steps in the basic PBIL algorithm. The object of the algorithm was described by Baluja and Caruana (1995, 5) in their seminal article:

> "...to create a real valued probability vector which, when sampled, reveals high evaluation solution vectors with high probability."

They clarified this with the example that, if a good solution to a problem can be encoded as a string of alternating 0's and 1's, then a suitable final probability vector would be 0.01, 0.99, 0.01, 0.99, etc. Baluja (1996) noted in a later article that in this way the algorithm attempts to *explicitly* maintain statistics about the search space that enables it to decide where to sample next.

In the case of the feature subset selection in this study, PBIL was used to find a binary mask that maximised the separability index (SI).

1 Initially the values of the probability vector were set to 0.5 for each bit corresponding to a potential feature. In this way a random sampling generator assigned a 0 or 1 with equal probability to any bit in the masks generated.

2 A number (N) of solution masks were then generated based on the probability vector. Each mask was evaluated using the SI.

3 The mask(s) with the highest values for SI were then selected.

4 The probability vector was then "pushed" towards the generated solution vector(s) with the highest SI evaluation. The distance that the probability vector was pushed was dependent on the learning rate parameter, α (Baluja & Caruana, 1995, 6).

In the equation in step (4) above:

- $p_{g-1}(x_i \mid D_{g-1})$ is the probability distribution of bit i in the old population,
- $p_{g-1}(x_i \mid D^{s}_{g-1})$ is the probability distribution of bit i among selected individuals, and
- α is the learning parameter referred to by Baluja & Caruana.

This equation adjusts the probabilities at each bit so as to increase the frequency of trial solutions which "resemble" the best solutions discovered in the earlier steps (Greene, 1996, 261).

5 Greene (1996) noted that PBIL can show a tendency to converge prematurely on a local optimum. Baluja recommended some form of random mutation of the probability vector in order to address this

issue.

In genetic algorithms, mutation maintains diversity by making it possible for a **0** or **1** to be randomly generated into a particular bit position. While this is not possible with PBIL (as bit value is determined through the probability vector), Greene suggested that the probability vector itself be altered in order to apply pressure on each bit's probability away from 0 or 1.

He called this the "forgetting" operator. It moves each probability bit slightly towards 0.5. This "forgetting" operator is denoted by a μ in the equation in step (5) in the figure above. Greene found that such a factor did result in slightly improved performance. This differed from the initial suggestion by Baluja & Caruana that such perturbation be made in a random direction to the probability vector.

6 Once the probability vector was updated as per steps (4) and (5), a new set of *N* masks was generated based on the new probability vector. The cycle continued for *g* iterations.

The values selected for the parameters to this search model are discussed later in this study.

17.7.3. Matlab Code for PBIL Used in this Study

The above algorithm was encoded and run using Matlab. The code used is included in Appendix A.2. Additional notes added to the code further describe the logical steps underlying the algorithm.

17.7.4. Results Achieved in Prior Research

Back et al (1996) found that, while using discriminant analysis, logit analysis and GAs to select predictor variables for bankruptcy prediction all result in the construction of different failure models, those constructed using GAs achieved superior predictive accuracy.

Although PBIL never appears to have been used in feature subset selection for failure prediction models, its structural resemblance to GAs and its comparative results indicate that it may well be a good algorithm for such an optimisation task.

Baluja and Caruana (1995) found that PBIL performed as well, or in certain instances better, than simple GAs on a number of standard optimisation tasks. In addition, Greene (1996) ran PBIL on a number of standard benchmark optimisation problems with results that consistently outperformed those of GAs.

PBIL has, however, been used for feature subset selection in other fields of study. Inza et al (1999) found that using PBIL for feature subset selection significantly improved the predictive accuracy of their Naïve-Bayes and ID3 models and considerably reduced the number of attributes in their "case study in the survival of cirrhotic patients with TIPS". The table below illustrates the improvement in their results:

	PBIL		No PBIL	
	Accuracy	*Number of attributes*	*Accuracy*	*Number of attributes*
Naïve-Bayes	86.53%	11	72.90%	77
ID3	85.97%	7	60.75%	11

Table 17.1. Comparative results of PBIL and non-PBIL approaches to feature selection in models constructed by Inza et al (1999, 12) for the prediction of survival of cirrhotic patients treated with TIPS.

17.7.5. ADVANTAGES AND DRAWBACKS OF PBIL

Greene (1996, 264) noted that not only is PBIL easy to intuitively understand and implement but that it is its simplicity in use that is important. He justified this through citing the following characteristics of the algorithm:

- only the best performing trial solution is used;
- the algorithm is representation tolerant; and
- there are few, uncritical control parameters that need to be selected in its implementation.

In addition, where recombination in GAs may change a viable trial solution into a non-feasible solution, these problems do not seem to arise with PBIL. Non-feasible trial solutions are simply not evaluated.

Dorsey et al (1994, 24) noted that premature convergence can occur with genetic algorithms if an early generation has a small number of trials that give much better values for the evaluation function. This danger is also prevalent in PBIL in a situation in which a few exceptional bitstrings totally dominate all subsequent probability vectors before the entire solution space has been adequately searched. In fact, PBIL is worse than a standard GA at maintaining diversity as it has just a single probability vector (Greene, 1996).

However, Baluja and Caruana (1995) pointed out that a parallel version of PBIL may be feasible in maintaining such diversity (see discussion on areas for further research). Greene (1996, 266) proposed a "multi-start PBIL search optimised for rapid convergence (large learning rate and zero forgetting factor) terminating the search whenever [a predetermined criteria is reached]". After using this method to rapidly identify local optima, a detailed PBIL search can then be performed across a domain limited to the size of the neighbourhood and centred on each of these points individually.

17.7.6. Areas for Further Research

Additional refinements that can be made to PBIL for potentially improved feature subset selection include:

- Baluja (1996) tested an alternative version of PBIL learning that not only pushes the probability vector towards the best solution, but also moves this vector away from the worst.
- Inza et al (1999) pushed the probability vector towards the best M solutions rather than the single best bitstring evaluated.
- Baluja (1996) proposed introducing parallel running PBIL algorithms that evaluate multiple probability vectors. A form of crossover can then also be introduced.
- Baluja and Caruana (1995, 8) proposed a variation on PBIL where the probability vector is incrementally updated as each new trial is generated rather than updating it only from a few of the best trial solutions generated after each new iteration.

CHAPTER 18
APPLICATION OF FEATURE SUBSET SELECTION METHODOLOGY

This chapter describes how the databases of company features constructed in Section B were used as inputs into the feature subset selection methodology described in the previous chapter.

The steps that were followed in the application of this methodology were as follows:

- Step 1: The necessary PBIL **parameters** were selected.
- Step 2: Then **preliminary PBIL tests** were run on the reduced "more obviously separable" data set.
- Step 3: Finally, PBIL was run on the full data sets in order to determine the **final feature subsets** for input into the classifiers (discussed in the following section).

The discussion in this chapter proceeds accordingly. However, the "reduced" and "full" data sets are discussed first.

18.1. DEFINING THE DATA SETS SUBJECT TO PBIL FEATURE SUBSET SELECTION

18.1.1. Defining Full Data Sets

Part of the research objective of this study is to construct a different model for corporate failure prediction for each of the three forecast periods:

- One year forward failure prediction (i.e. predicting failure based on the last set of financial statements published by a company prior to its defined date of failure).

- Two year forward failure prediction (i.e. predicting failure based on the company financial information available two financial year ends prior to its defined date of failure).

- Three year forward failure prediction (i.e. predicting failure based on the company financial information available three financial year ends prior to its defined date of failure).

Financial data was collected for each failed company (and corresponding paired non-failed company) for a maximum of three years and minimum of two years prior to the said company's date of failure. As noted in the previous section, a separate *3-year dbase* and *2-year dbase* was constructed for those companies with three years and two years of available data, respectively (note that all companies falling within the *3-year dbase* would also have been included in the *2-year dbase* but not vice versa).

Therefore, each one and two year forward prediction model could have been constructed using a different number of years of available data. For example, the one year forward prediction model could have been constructed using one, two or three years of data prior to the defined date of failure.

As a result, the PBIL feature subset selection procedure was run across all possible combinations of forward prediction data sets in order to identify which data sets were best able to forecast failure one, two and three years forward. These possible combinations (and their reference names for the discussion that follows) are presented in the table below:

	Using 1 year of data prior to date on which prediction is made	Using 2 years of data prior to date on which prediction is made	Using 3 years of data prior to date on which prediction is made
1-Year Forward Prediction Model	**1-yr fwd/1-yr data model**	**1-yr fwd/2-yr data model**	**1-yr fwd/3-yr data model**
2-Year Forward Prediction Model	**2-yr fwd/1-yr data model**	**2-yr fwd/2yr data model**	Insufficient Data
3-Year Forward Prediction Model	**3-yr fwd/1-yr data model**	Insufficient Data	Insufficient Data

Table 18.1. The various one, two and three year forward forecasting data sets that were subjected to PBIL feature subset selection.

18.1.2. Defining the Reduced Data Sets

In addition, as noted in the previous section, additional "reduced sample" databases, in which the borderline failure and non-failure cases were eliminated, were created.

Initially, only these databases were tested. The rationale behind these preliminary tests was as follows: if PBIL could not be used to select a subset of features with a significant separability index on the reduced sample, then such a procedure on the full sample would be of little value.

18.1.3. Importing the Data Sets into Matlab

All these data sets were imported as matrices into Matlab using the Matlab import wizard. Each matrix had the company input vectors in rows and the features in columns (see Figure 16.3.). Target labels (-1 for failure and +1 for non-failure) were appended onto the end of each input vector.

18.2. STEP 1: SELECTION OF PBIL PARAMETERS

The parameters that need to be selected in order to run the PBIL algorithm are non-critical (Greene, 1996). It was outside of the scope of this study to perform detailed research in order to find these optimal values. As a result, the methods for determining the parameter values in this study were crude and involved subjectivity on the part of the researcher. More detailed investigations in this area are suggested as an area for further research in Chapter 23.

The parameters of the PBIL algorithm include:

- Number of trials (*N*) and number of epochs (*g*)
- Learning parameter (α)
- Forgetting operator (μ)

18.2.1. Number of Epochs (*g*) and Trials (*N*)

N is the number of random (as guided by the probability vector) masks generated on each iteration. The "pbilmask" Matlab code in Appendix A.2. refers to the number of trials as the variable "*ntrials*" and the number of iterations as "*epochs*".

A number of runs of the PBIL algorithm were performed. On each run, different values and combinations of *g* and *N* were tested.

Through a heuristic search, the following ranges were determined as optimal for each of these parameters:

- *N*: 1000 to 7500
- *g*: 75 to 125

It was noted that as the value for *N* increased, the number of iterations that PBIL took to settle on an optimal feature subset decreased (i.e. the required value for *g* decreased). In other words, as the number of trials generated by each epoch increased, so the algorithm was more rapidly able to identify the optimal feature subset.

The results presented below disclose the parameters that were used in generating those results.

18.2.2. LEARNING PARAMETER

Greene (1996, 264) noted that an alpha of 0.1 performs satisfactorily across a wide variety of optimisation tasks. A superficial assessment, using a number of alpha values, confirmed that other alpha values did not yield significantly different results.

As a result, this parameter was fixed at 0.1 for all PBIL runs.

18.2.3. FORGETTING OPERATOR

Greene also noted that a forgetting factor of 0.005 performed satisfactorily across a variety of optimisation tasks. Once again, a superficial assessment of this parameter did not indicate that a different value would produce significantly better results.

18.3. STEP 2: PRELIMINARY PBIL RESULTS - REDUCED DATA SETS

The PBIL feature subset selection algorithm (Appendix A.2.) was run on a *3-yr fwd/1-yr reduced data* feature set, a *2-yr fwd/2-yr reduced data* feature set and a *1-yr fwd/3-yr reduced data* feature set, each from the reduced sample databases discussed above. These selected reduced data sets used the maximum number of years of data available for a particular forecast period.

The algorithm was run three times on each data set. On each run, g was set to 100 and N was set to 250, 500, and 750, respectively. These values for N were small in relation to the optimal values identified in the detailed testing on the full data set. This was in order to save computing time as these tests were only intended to serve as an indication of the potential for full data set separability.

	Forward Forecasting Data Sets (using maximum data)		
	3-yr fwd	2-yr fwd	1-yr fwd
SI Value	87.78%	92.22%	95.56%
N	500	750	750

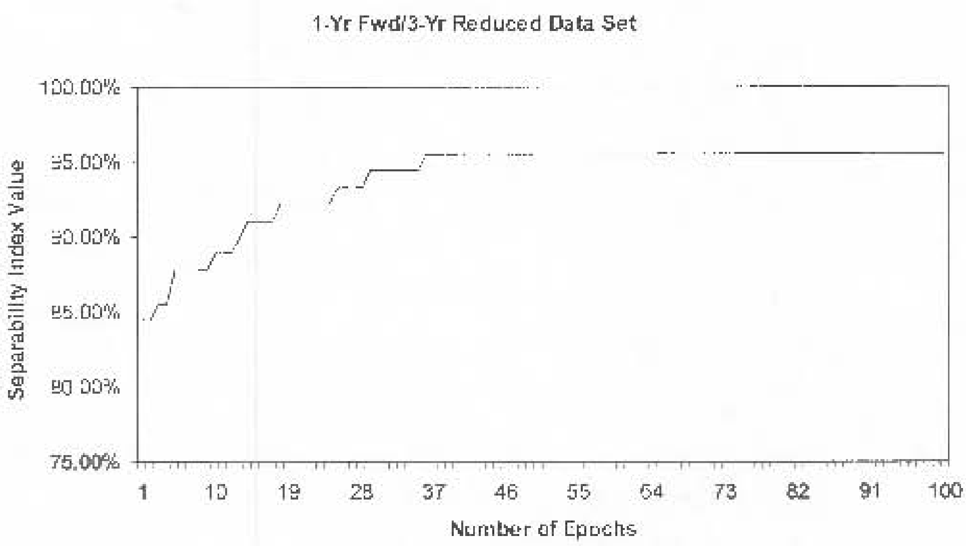
Table 18.2. Results of the best performing preliminary PBIL feature subset selections on the reduced data sets.

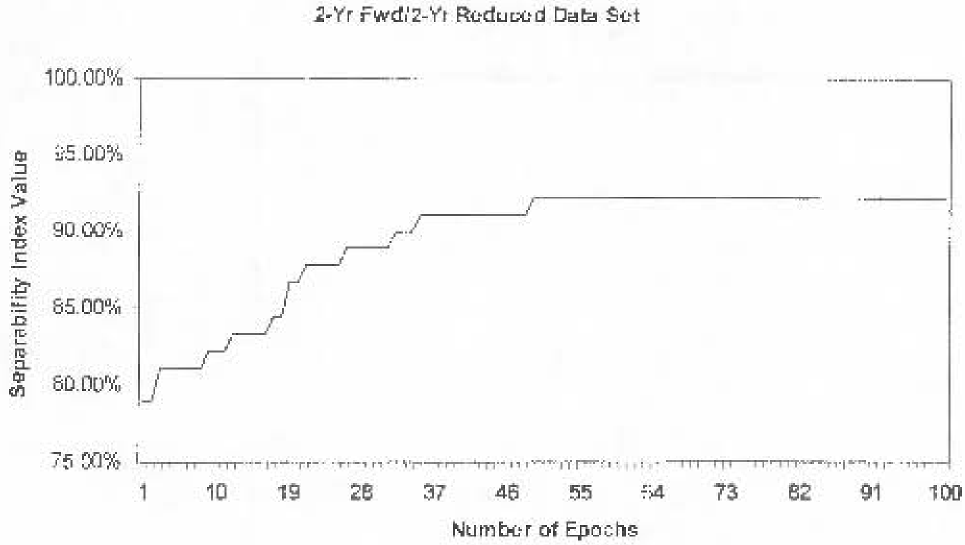
These results were consistent with the expectation that separability would decrease with an increase in the forward forecast period. Relative to the prediction accuracies attained by other studies in the corporate failure literature, the separabilities disclosed in Table 18.2. were considered to be of sufficient significance to proceed to the next stage of testing.

Figure 18.1. graphs the best SI value for each of the hundred epochs generated in attaining each of the results presented in Table 18.2. Note that the PBIL algorithm:

- converged on an optimum SI value after numerous epochs, and
- remained constant at the best achieved level.

This was an indication that the PBIL algorithm was functioning as intended.





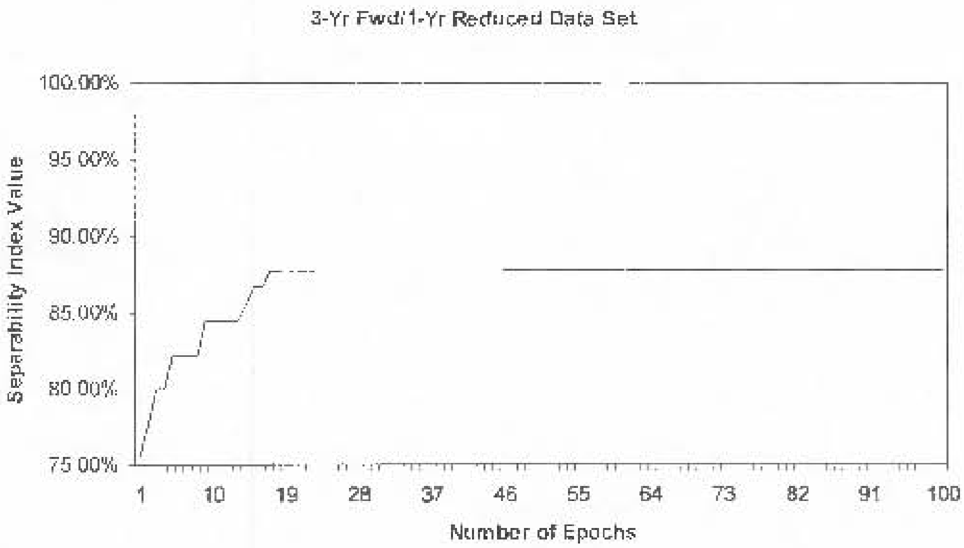


Figure 18.1. Improvement of the separability index values across epochs for the PBIL feature subset selection run on the reduced data sample sets

18.4. STEP 3: FINAL FEATURE SUBSET SELECTION - FULL DATA SETS

The following procedures were followed in the selection of the final feature subsets:

- The different data sets (Table 18.1.) were tested in order to determine which data set was optimal for each of the three forecast periods.
- Further PBIL testing was then performed on each of the optimal data sets.
- Finally, the subset(s) of features that best "separated" the failed and non-failed classes were selected.

As noted earlier, through a simple heuristic search, it was determined that the optimal range of values for *N* and *g* were somewhere between 1000 to 7500 and 75 to 125, respectively.

18.4.1. Selection of the Optimal Data Set for each Forecast Period

Each of the six data sets described in Table 18.1. were subjected to five PBIL feature subset selection runs using the follow values for *N* and *g*:

- *N* = 1 000, *g* = 125;
- *N* = 2 500, *g* = 125;
- *N* = 4 000, *g* = 100;
- *N* = 5 500, *g* = 75;
- *N* = 7 500, *g* = 75;

The number of PBIL runs for each data set was limited to five as a result of the required computing time. As an example, approximately 4 hours on a Pentium III 850 MHz computer was required in order to run the "pbilmask" algorithm (Appendix A.2.) on the *1-yr fwd/3-yr data* set (i.e. the largest data set), with *N* set to 7500 and *g* set to 75. These five runs, however, were deemed to be sufficient to test each data set.

The results for the best runs for each data set are presented below. The "SI prior to PBIL" column gives the SI value for the full feature set (i.e. before any subset had been selected).

	SI prior to PBIL	SI	*N*	*g*
1-Year Forward Forecasting				
1-Yr Fwd/1-Yr Data	63.41%	79.268%	2 500	125
1-Yr Fwd/2-Yr Data	65.85%	89.024%	5 500	75
1-Yr Fwd/3-Yr Data	66.39%	90.984%	5 500	100
2-Year Forward Forecasting				
2-Yr Fwd/1-Yr Data	62.80%	86.585%	4 000	100
2-Yr Fwd/2-Yr Data	58.20%	87.705%	4 000	75
3-Year Forward Forecasting				
3-Yr Fwd/1-Yr Data	60.66%	90.164%	5 500	100

Table 18.3. The best result for the five PBIL feature subset selection runs on each of the six different data sets.

These results show that the data sets that included the maximum possible number of years of data for each forward forecast period were best at separating the failed and non-failed companies. For example, the best separability obtained for one year forward forecasting was 90.164% on the data set that included all three years of data prior to the defined date of failure. This was more than a 23% improvement in separability on the data set prior to feature subset selection (as measured by the SI value). A similar result was obtained for two year forward forecasting.

The objective of this study was to construct the best possible model for each forecasting period. As a result, the most "separable" data sets for each forecast period were selected and taken into the forecasting models (described in the following section). The remaining data sets were discarded.

Thus, the prediction models were constructed using the following data sets:

- one year forward forecast model: *1-yr fwd/3yr data*
- two year forward forecast model: *2-yr fwd/2-yr data*
- three year forward forecast model: *3-yr fwd/1-yr data*

18.4.2. Preliminary Analysis of Feature Subsets

Five different feature subset binary masks were constructed for each data set in the process of determining the optimal data set described above. These binary masks for each optimal data set selected, were then assessed for similarity.

(a) Assessment of similarity between different feature subset binary masks

Hamming distance was used to perform this analysis. Hamming distance is calculated by counting the number of bit positions in which two feature subset binary masks disagree.

It was found that, within each data set, the masks that were generated differed significantly from each other despite having near-similar separability. There were cases in which two binary masks differed in more than 50% of their bit positions but still had equal separability index values.

In other words, the optimal feature subsets discovered on each PBIL run varied significantly in terms of the actual features selected but not in terms of separability.

(b) Generation of additional dissimilar feature subset binary masks

Additional PBIL runs were then performed.

Firstly, the Matlab code was adjusted so as to save any mask in a single PBIL run with an SI value greater than a specified floor value.

However, the Hamming distances between those masks collected on a single PBIL run were small (i.e. the masks were all very similar, except for a few bit positions). This was possibly because the algorithm focuses on a single "area" of the search space during each PBIL run. As a result, all masks with high separability index values produced on a **single** PBIL run were not significantly dissimilar.

However, as **independent** PBIL runs were able to produce dissimilar masks of equal separability, the PBIL algorithm (through its random bitstring generations) was able to focus on different areas of the search space on each of these different runs. The fact that there were dissimilar masks of equal separability indicated that there were numerous optimal areas to the solution space.

In Chapter Five it has been noted that the interpretation of the information content of a financial data item is dependent on with which other variable(s) that item is analysed. The various dissimilar feature subsets selected on independent PBIL runs may be as a result of this characteristic of financial information. The variations in these subsets is analysed for reasonability below.

Therefore, instead of collecting numerous masks on a single PBIL run, an additional five independent PBIL runs, with differet values for the PBIL parameters, were performed on each of the selected data sets. The masks produced were then saved.

18.4.3. Final Feature Subsets

Based on the work performed up until this point, there were ten binary masks from which to choose the final feature subset for each forecast period (five from the initial runs and five from the additional runs). All these masks, with their respective Separability Index values, are included in Appendix H

The best binary masks (or "optimal feature subsets") were selected based on which had achieved the highest Separability Index value. Using this as the selection criterion, one optimal feature subset was selected for the one year forward forecast model. However, two optimal feature subsets, each with identical SI values, were selected for the two year forward forecast model. Similarly, three optimal feature subsets were selected for the three year forward forecast model. All the selected subsets have been identified in Appendix H.

The following table describes each of the optimal feature subsets. For the remainder of this report, each optimal feature subset will be referred to as it is named in this table.

	SI Value	*N*	*g*	Label per App H	Number (%) of Features
1-Yr Fwd Model					
- Optimal Feature Subset A	90.984%	5 500	100	A1	72 (44%)
2-Yr Fwd Model					
- Optimal Feature Subset A	87.705%	4 000	75	B1	40 (34%)
- Optimal Feature Subset B	87.705%	5 000	80	B5	56 (48%)
3-Yr Fwd Model					
- Optimal Feature Subset A	90.164%	5 500	100	C5	17 (29%)
- Optimal Feature Subset B	90.164%	3 000	125	C6	16 (27%)
- Optimal Feature Subset C	90.164%	7 500	75	C10	18 (31%)

Table 18.4. Details of the final feature subset(s) selected for each forecast period

The table above discloses the number of features, as well as the percentage of total features, contained in each subset. It was encouraging to note that as little as 27% of available features were able to obtain an SI value of over 90% for the three year forward prediction model. The highest proportion of features contained in any subset was 48% for Optimal Feature Subset B of the two year forward prediction model. It, therefore, appears as if PBIL performed well in identifying a smaller subset of features with improved data separability.

The graphs depicting the improvement in the separability index across epochs for each of the PBIL runs are presented below (only the graphs for each forecast period's Optimal Feature Subset A is presented).

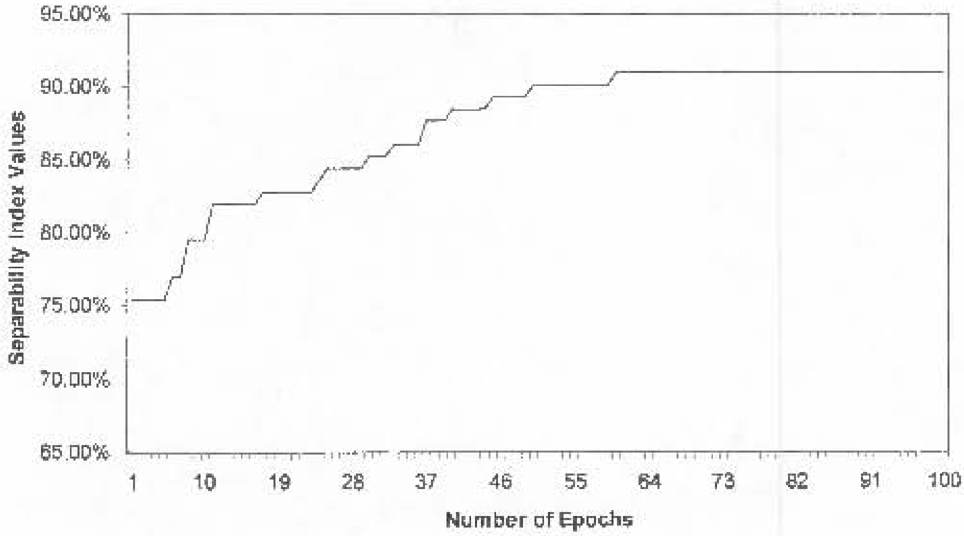


Figure 18.2. Graph of the improvement in the *1-yr fwd/3-yr data* set separability index across epochs for the PBIL run that produced Optimal Feature Subset A.

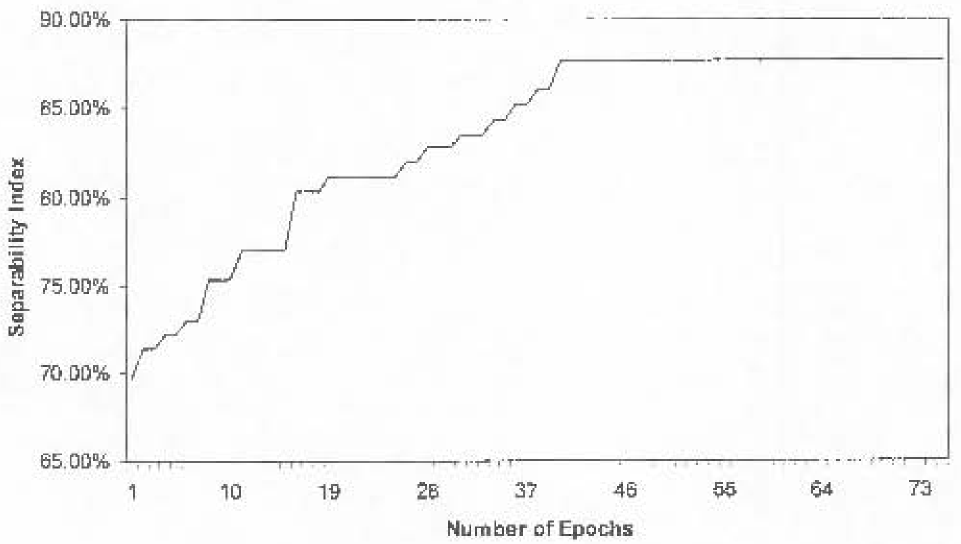


Figure 18.3. Graph of the improvement in the *2-yr fwd/2-yr data* set separability index across epochs for the PBIL run that produced Optimal Feature Subset A.

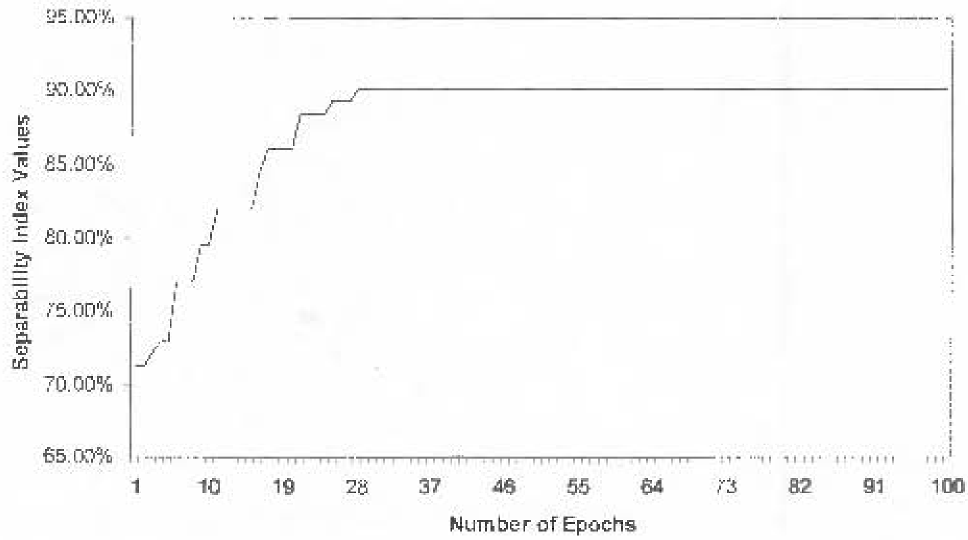


Figure 18.4. Graph of the improvement in the *3-yr fwd/1-yr data* set separability index across epochs for the PBIL run that produced Optimal Feature Subset A.

A brief analysis of the similarity between the equally-separable feature subsets of the two and three year forward forecast models is disclosed in the tables below. The method for calculating Hamming distance has been discussed above.

Optimal Feature Subset	A			B		
	Ham Dist	% Tot Feat	% Sel Feat	Ham Dist	% Tot Feat	% Sel Feat
A				46	39.3%	82.1%
B	46	39.3%	115%			

Table 18.5. Analysis of similarity between the Optimal Feature Subsets for the two year forward forecast model (Hamming distance, Hamming distance as a % of total available features, Hamming distance as a % of selected features)

Optimal Feature Subset	A			B			C		
	Ham Dist	% Tot Feat	% Sel Feat	Ham Dist	% Tot Feat	% Sel Feat	Ham Dist	% Tot Feat	% Sel Feat
A				7	11.9%	43.8%	7	11.9%	38.9%
B	7	11.9%	41.2%				4	6.8%	22.2%
C	7	11.9%	41.2%	4	6.8%	25.0%			

Table 18.6. Analysis of similarity between the Optimal Feature Subsets for the three year forward forecast model (Hamming distance, Hamming distance as a % of total available features, Hamming distance as a % of selected features)

The similarity between Optimal Feature Subsets A, B and C of the three year forward model is greater than the similarity between Optimal Feature Subsets A and B of the two year model. However, the significant difference between each subset, especially when based on the number of features selected, is an indication that each subset may in fact represent a different area of the solution space as searched by the PBIL algorithm.

18.5. ECONOMIC INTERPRETATION OF SELECTED FEATURE SUBSETS

Once the final feature subsets had been selected, a basic analysis of the features was performed. This was to determine if the models to be constructed would have a sensible economic interpretation. If the selected features could not have been interpreted as part of a plausible economic model then there may have been a spurious relationship between the selected features and data separability.

18.5.1. CASH-FLOW CONCEPT INTERPRETATION

As described in Chapter Fifteen, Beaver's "Cash-Flow Concept" (1966) was used in the process of selecting the full set of potential features to be input into the feature subset selection procedure. This concept describes the possibility of failure of the firm in terms of the possibility of the "reservoir" representing the firm's assets running dry. Figure 15.1. describes this concept..

The full set of potential features were also categorised into "information groupings" (e.g. liquidity, operating efficiency, etc.). Each of these groupings could then be viewed as measuring the various components of the "Cash Flow Concept" as illustrated in Table 18.7. below.

In analysing the selected feature subsets, it was noted that ALL the feature subsets included at least one feature describing each of these information groupings for EVERY year of data to be included in the model, except (Appendix H.4. presents the six optimal feature subsets and summarises the frequency of feature inclusion):

- None of the 3 Yr Fwd model's optimal feature subsets included any liquidity ratios.
- The 1 Yr Fwd model's Optimal Feature Subset A did not include a measure of operating efficiency for 3 years prior to failure.

This is consistent with the "Cash-Flow Concept" - one would expect that all components impacting the reservoir would need to be measured by a failure prediction model.

In fact, in both these cases in which a ratio from one of the groupings was not present in a year of data, other data items measuring that cash-flow component (as described in the table below) were still selected for the feature subset.

"Cash-Flow" Concept Component	Variable Grouping
Size of reservoir	(L) Liquidity (M) Financial market (G) Size and growth
Size of liquid-asset inflows	(L) Liquidity (E) Operating efficiency (C) Cash flows
Size of debt	(S) Solvency (M) Financial market
Size of expenditures from operations	(L) Liquidity (P) Operating profitability (C) Cash flows
Variability of earnings and claims	(R) Risk analysis

Table 18.7. Summary of "cash-flow" concept components addressed by each variable group.

18.5.2. Most and Least Commonly Selected Features

It was possible for each feature to have been selected a maximum of ten times across the various forecast periods and optimal feature subsets:

In the 1 Yr Fwd model, any feature could have been selected three times in the 1 Yr Fwd Optimal Feature Subset A, as this model included data for all three years prior to failure. Similarly, in the 2 Yr Fwd models, Optimal Feature Subsets A or B could have included any variable twice, respectively. The three optimal feature subsets identified for the 3 Yr Fwd models could each have included a particular feature once.

Most commonly selected features overall

Table 18.8 below lists the most commonly selected features. These features were selected six to eight times across the various models and years.

Feature Code	Feature	Number of Inclusions	Cash-Flow Component Addressed
M2	Total dividend yield	8	Liquid-asset outflows
L9	Payables turnover	7	Liquid-asset outflows & debt
P1	Operating profit margin	7	Liquid-asset inflows & outflows
S1	Debt equity ratio	7	Size of debt
S7	Debt & cont. to assets	7	Size of debt
C7	Cash flows from ops to debt	7	Liquid-asset inflows & outflows
P2	Net profit margin	6	Liquid-asset inflows & outflows
P5	Return on ordinary equity	6	Liquid-asset inflows & outflows
S3	Int-bearing debt to assets	6	Liquid-asset outflows & debt
G2	Log (market capitalisation)	6	Size of reservoir
G4	Retention ratio	6	Size of reservoir & inflows

Table 18.8. Features most commonly selected across models' optimal feature subsets and years prior to failure.

An analysis of this subset shows that this sub-group of ratios also measures all of the components of the "Cash-Flow Concept". In fact, all the information groupings are represented within this subset, except:

- operating efficiency (possibly because there are only 3 features from which to select) and
- risk analysis (because these ratios are only available for selection in the 1 Yr Fwd data set).

The fact that these most commonly selected features cover the spectrum of analysis of a company is encouraging because it means that the feature subset selection procedure did not become trapped in a specific area of the search space. Importantly, this makes intuitive economic sense.

Liquidity ratios (L)

The most commonly selected liquidity ratio was payables turnover (selected seven times). This ratio measures the critical outflows relating to the payment of creditors. It also feeds into the cash management cycle. This ratio has been identified as a key indicator of failure in many corporate failure studies (Appendix B).

The non-group current ratio and non-group quick ratio were the least commonly selected ratios (once and twice respectively). This is an indication that perhaps the intra-group short-term balances are not useful in distinguishing between failure and non-failure.

It was interesting to note that liquidity ratios were not selected at all for the 3 Yr Fwd model, but then increasingly more often for the 2 Yr Fwd model (on average more than three times per year), and then, finally, significantly often for the 1 Yr Fwd model (nearly six times per year on average). This may be an indication that liquidity ratios carry information regarding the impending failure of a company as liquidity becomes worse closer to the death.

Operating efficiency ratios (E)

Each of the three potential operating efficiency ratios E1, E2 and E3 were selected across feature subsets and years three, four and four times, respectively. While no measure of operating efficiency was clearly favoured over another, a measure of operating efficiency was included in ALL years of feature subset data AT LEAST once (with the exception of the third year prior to failure for the 1 Yr Fwd model's Optimal Feature Subset A).

Therefore, operating efficiency was interpreted as being useful in the forecast of failure. However, the information content of total asset turnover, fixed asset turnover and equity turnover is relatively equal. This makes intuitive sense if one considers that the different sizes measured in the denominator of each of these ratios are effectively measured by the "size and growth" variable group, leaving turnover as the common but critical information input.

Operating profitability ratios (P)

Operating profit margin, net profit margin and return on ordinary equity were the most commonly selected operating profitability measures (seven, six and six times, respectively). These measures have been commonly used in the corporate failure literature (Appendix B). They provide information on profit margins before and after interest and tax and also on how that profit relates to a measure of return on investment.

Interestingly, return on total equity and return on total capital were never included simultaneously in the same year of data. This reflects the understanding that these measures carry similar information content.

Solvency ratios (S)

This information group was the most commonly included group. No year of feature subset data included any fewer than three of these ratios. This was a good reflection of the definition for corporate failure used in this study – where solvency impacts the winding-up of a company (see Chapter Three), where the definition includes the inability to meet debt repayments and where a firm with negative net worth is deemed to have failed (see Chapter Thirteen).

The only solvency ratio not included any number of times was the total commitments to asset ratio (not included in any feature subset). This was to be expected if one considers the degree to which such "commitments" accounting disclosures are subject to management discretion and manipulation.

Cash flow ratios (C)
The most commonly selected cash flow ratio was cash flow from operations to total debt (included seven times).

The ratios selected least for use in the forecast models were cash flow to total debt (selected once) and cash flow to interest-bearing debt (not selected at all). This made intuitive sense as these ratios will include effectively similar information to cash flow from operations to total debt. The fact that cash flows before accounting for tax, interest and dividends were preferred over those cash flows after such deductions was an interesting observation. However, this information had been already included in the numerous selections of net profit margin and operating profit margin.

The proportion of dividends that was not distributed in cash was also not selected for any feature subset. This may, once again, be as a result of the poor and inconsistent disclosure of such information in company financial statements.

Market ratios (M)
Total dividend yield was selected eight times for use across various feature subsets. Clearly, dividends return on market share prices are an important indicator with regards to corporate failure prediction. While other features were all included numerous times, dividend yield eclipsed all other measures of market value performance. It is consistent with the literature on corporate failure that dividends carry important information (see Chapter Thirteen).

The least included market ratio was trading turnover (only included twice).

Risk analysis ratios (R)
Of all the potential risk analysis ratios, only sales variability over three years was used by any single feature subset.

This was unexpected, as risk analysis is considered to be critical in the assessment of the probability of failure. The low use of this information group may be because there were an insufficient number of years of data over which to calculate variability, or because variability was captured in other measures (such as the inclusion of net profit margins and turnover ratios over multiple years), or because the wrong risk analysis ratios were included in the full feature set.

Size and growth ratios (G)

The logarithm of market capitalisation and the retention ratio were the most commonly selected ratios from this information group (six times each). The former is a measure of size and the latter a measure of growth – this is encouraging as it is unlikely that they overlap on information content, despite being from the same information group.

The logarithm of firm assets was included only twice. On both these occasions, the logarithm of market capitalisation was not included in the feature subset. This is intuitively correct as both these features will convey very similar information about the size of the firm concerned.

Non-financial ratios (N)

A spread of these ratios were selected across subsets and years prior to failure. This is consistent with the fact that each of these measures contains relatively unique information.

18.5.3. Economic Analysis of Dissimilarities between Selected Feature Subsets for each Model

As noted above and in Chapter Five, the interpretation of the information content of a financial data item is dependent on with which other variable(s) that item is analysed. Therefore, the set of potential features selected for this study was chosen ignoring any information that may have been double-counted in their selection. It was recognised that, while many features may be highly correlated, there may be unique information contained in each that may provide improved explanatory power when viewed in combination with other features.

The various dissimilar feature subsets selected on independent PBIL runs may be as a result of this characteristic of financial information, combined with the ability of PBIL to search different areas of the solution space through its randomised bitstrings.

2 Yr Fwd Model (Optimal Feature Subsets A and B)

As per Table 18.5., the Hamming Distance of 46 between Optimal Feature Subsets A and B for the 2 Yr Fwd model means that nearly 40% of the features selected in proportion to the total available features for the two subsets are different.

An economic analysis of the differences reveals the following:

- While each subset does not include the same current/quick ratio, each does include one of L1 to L5. As these can be deemed to carry similar information, these feature differences are not deemed to be significant (Hamming Distance = 2).

- While Optimal Feature Subset B includes all the components of the cash conversion cycle (L6 to L9) for two years prior to failure and Subset A does not, Subset A includes the cash conversion cycle ratio itself (L10) where B does not. These effectively measure the same information (Hamming Distance = 4).

- While Optimal Feature Subset B does not include total asset turnover two years prior to failure (where Subset A does), Subset B then includes the size variable Log(Assets). In this way the asset size of the company has still been accounted for (Hamming Distance = 2).

- While each subset does not include the same measure of profit margin (P1 and P2), each does include at least one measure of profit margin (P1 or P2) (Hamming Distance = 4).

- In terms of measuring solvency, Optimal Feature Subset B two years prior to failure includes measures of total debt S6, S7 and S9. Optimal Feature Subset A does not include these features, but rather includes measures of long-term debt (S2, S3, C4 and C8). The difference between long-term debt and total debt is that total debt includes the non-interesting bearing current liabilities. However, where Subset B lacks explicit information with regards to what portion of the debt bears interest, it also includes the fixed charge coverage ratio (C2) that measures the interest burden, where Subset A does not (Hamming Distance = 9).

- Optimal Feature Subset A includes the log(assets) size measure where Subset B rather includes the log(market capitalisation) for three years prior to failure (Hamming Distance = 2).

- Optimal Feature Subset B includes retention ratio (G4) and growth rate (G5) three years prior to failure, where Subset A does not. However, Subset A includes return on ordinary equity (P5) for this year – an inverse version of the growth rate (G5) (Hamming Distance = 3).

This preliminary analysis helped to explain 26 of the 46 differences between the two feature subsets. This indicated that the differences between the two subsets with regards to information content were not as extreme as the initial Hamming Distance analysis indicated.

3 Yr Fwd Model (Optimal Feature Subsets A, B and C)

As per Table 18.6., the Hamming Distances of 7 and 4 between the various feature subsets for the 3 Yr Fwd model means that only between 7% and 12% of the features selected in proportion to the total available features for the subsets are different. This is significantly more stable than the case of the 2 Yr Fwd Model.

A brief economic analysis of these differences revealed the following:

- Optimal Feature Subset A includes return on total equity (P4) while Subsets B and C include return on ordinary equity (P5) (Hamming Distance = 2)

- Optimal Feature Subset B includes a measure relating historic cost accounting to market value in the form of market value of equity to total debt (M4). Subset A includes price book ratio instead (M5) (Hamming Distance = 2).

18.6. IN SUMMARY: END RESULT OF FEATURE SUBSET SELECTION PROCESS

As a conclusion to the feature subset selection process:

Firstly a reduced data set was used as a preliminary test to see if PBIL and the available pre-processed features were obtain to obtain significant data separability.

Then six data sets, covering all three forecast periods, (see Table 18.1.) were subjected to five PBIL runs using parameters selected on a trial-and-error basis. A single optimal data set was selected for each forecast period and input into the PBIL algorithm for an additional five runs.

The resulting ten binary masks, each representing a different subset of features, were assessed for separability using the Separability Index (see Appendix H). Those with the highest SI value were selected for final input into the classifier. The construction of the classifier is discussed in the following section.

The feature subsets were then subjected to a preliminary analysis in order to test their economic feasibility.

SECTION D
EMPIRICAL MODEL CONSTRUCTION: CLASSIFIER CONSTRUCTION AND EVALUATION

CHAPTER 19

CORPORATE FAILURE PREDICTION MODEL: K-NEAREST NEIGHBOUR CLASSIFIER

In this section, the optimal feature subset(s) for each forecast period are input into a k-Nearest Neighbour (this chapter) and Kernel Ridge Regression (next chapter) classifier. Each chapter provides an explanation, as well as a justification for the use, of the specific technique employed.

The models are then evaluated, taking into consideration the different types of errors and their relative costs.

19.1. DISCUSSION OF THE K-NEAREST NEIGHBOUR (KNN) CLASSIFIER

The most basic proximity-based classifier involves classifying an unknown test case with the same classification label as that of the most similar known training case. This is known as the nearest neighbour classifier.

19.1.1. EXPLANATION OF THE FUNCTIONING OF THE NEAREST NEIGHBOUR CLASSIFIER

In order to find the nearest neighbour to a particular case, the similarity between different cases needs to be measured. The relative similarity between two cases can be measured by calculating the Euclidean distance between them in N-dimensional space (where N is the number of features of each input vector). In this way, a test point in $\Re^N$ is assigned the classification of the **closest** training point to it.

The general formula for the calculation of Euclidean distance is:

$$dist = (\sum_{n=1}^{N}(a_n - b_n)^2)^{0.5} \tag{19.1.}$$

where a and b are two points in N dimensional space and a_n and b_n are the n^{th} features by which both points a and b are defined.

Graphically, the nearest neighbour can be illustrated as follows:

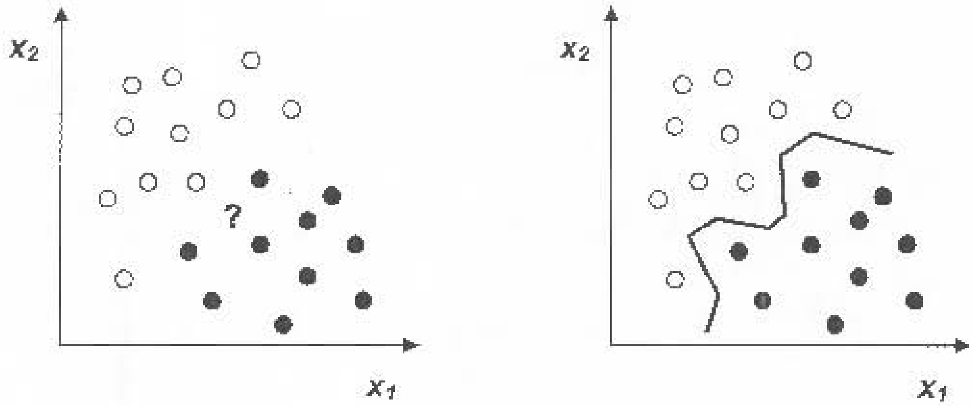


Figure 19.1. Graphical illustration of nearest neighbour classifier.

The query point in the left hand graph would be classified as "black" as it is closest to the black dots. The graph on the right illustrates the decision boundary produced by such a classifier. Bishop (1995, 58) noted that each segment of such a boundary will be the perpendicular bisector between the closest two data points.

19.1.2. Averaging the Classification Result over k Nearest Neighbours

The disadvantages of the basic nearest neighbour classifier are:

- its **sensitivity to outliers** (a test point situated within a cluster of a particular class may be misclassified as a result of lying nearest to an outlier of a different class), and
- its **tendency to "overfit"** and generalise poorly in the situation in which classes are not clearly separated (the nearest neighbour of a point in an intermingled data set may be of little use in generalising the classification of such point).

In such situations, the decision boundary (as illustrated in Figure 19.1. above) may become "convoluted and irregular" (Greene, 2002, 2).

Greene noted that such problems can be mitigated in two ways:

- by assigning an observation to a group to which the majority of its k nearest neighbours belong (kNN classifier), and/or
- by diffusing the region of influence of each training point through the use of radial basis functions (for example, kernel ridge regression, which is developed in the following chapter).

19.1.3. THE USE OF KNN IN PRIOR RESEARCH

> "Despite its simplicity, the nearest neighbour classifier performs rather well across a wide range of real-world problems, yielding results close to the state of the art on many of them." (Greene, 2002, 1)

KNN is a simple classifier that has been applied across disciplines in the literature (see Payne, Edwards & Green, 1995 for a summary). Tam & Kiang (1992) applied this methodology specifically to the task of bank failure prediction. They compared the kNN to linear, logistic, neural network and ID3 classifiers.

Their results showed that, while the 1NN and 3NN classifiers had the lowest accuracy, these models were still able to achieve comparable results (presented in Table 19.1.). The performances of the nearest neighbour classifiers were still competitive in relation to the performances achieved on most other models.

	Misclassification rates using the holdout sample	
	One-year Prior	Two-years Prior
DA	15.9%	17.5%
Logit	18.2%	7.5%
1NN	22.8%	22.5%
3NN	22.8%	22.5%
ID3	20.5%	20.0%
Net1	18.2%	22.5%
Net2	14.8%	16.3%

Table 19.1. Misclassification rates of the various models using the hold-out sample (reproduced from Tam & Kiang, 1992, 940)

However, the performance of kNN in the above study aside, the model has been applied with great success on other classification problems.

For example, in a study performed by Bryant (1997), a case-based reasoning (CBR) technique significantly outperformed a comparative logit model in the task of predicting corporate failure. The CBR used an adaptation of the nearest neighbour classifier in order to index the cases collected in the study.

19.2. JUSTIFICATION FOR THE USE OF *K*NN IN THIS STUDY

19.2.1. GENERAL ADVANTAGES OF kNN

The kNN classifier has been shown to be successfully applied across a wide range of classification problems. Despite its relatively poor performance in the study by Tam & Kiang (1992) discussed above, its success in Bryant's study and in other areas of research indicates that it may be suitable for this study.

In addition, it has the advantage of being computationally less demanding than other proximity-based techniques, as well as much easier to intuitively understand. In this way, kNN was an ideal intermediary model to employ before implementing the distance-weighted proximity based induction algorithms used in this study.

19.2.2. APPLICABILITY OF kNN TO THIS STUDY

As discussed in Chapter Seventeen, Thornton's separability index is similar to measuring the asymptotic result of a large number of train/test cycles with random splits using a nearest neighbour classifier (Greene, 2002, 2).

In the previous chapter, it was described how population-based incremental learning (PBIL) was employed in feature subset selection. The subset that maximised the separability between the failed and non-failed classes was deemed to be the "optimal feature subset". This was evaluated using the separability index (SI). The SI values for each of the optimal feature subsets for each prediction model have been presented in Table 18.4. in the previous chapter.

Later in this chapter it is shown that the SI value is identical to the classification accuracy rate of a 1NN using leave-one-out validation. Therefore, the high SI values of the optimal feature subsets were an indication that a 1NN could achieve a classification accuracy comparable to the best accuracy rates achieved in other studies published in the literature.

The use of kNN was further motivated by the fact that different values for k may further improve this classification accuracy.

19.3. THE USE OF CROSS-VALIDATION IN TRAINING AND TESTING THE MODEL

The goal for the construction of a failure prediction model is to find the feature subset and induction algorithm that together perform best on data previously unseen to the model.

19.3.1. THE RISK OF OVER-FITTING THE MODEL

A classic problem in failure prediction research, and indeed in all classification model construction, is that of "over-fitting". As Zhang et al (1999, 25) noted, when a model is trained on a set of data, the model will be tailored to fit that sub-sample and will often overly optimistically estimate the true error rate on that sub-sample. Generalisation and over-fitting have been discussed in detail in Chapter Nine.

19.3.2. APPLICABILITY OF LEAVE-ONE-OUT (LOO) VALIDATION TO THIS STUDY

The risk of underestimating the error rates when using kNN and KRR can be mitigated by using the validation techniques that have been laid out in detail in Chapter Ten.

There are, in total, only 164 companies in this study's sample (i.e. including failed and non-failed sub-samples). Therefore, due to the small sample size, a holdout-sample approach, where a portion of the data is set aside to validate the classifier, was not feasible. Hence, a cross-validation method was preferred.

In order to maximise the data used to train the classifiers, Leave-One-Out (LOO) validation was used. This has also been discussed in detail in Chapter Ten. The two major drawbacks of this approach, noted in Chapter Ten, are:

- **Computationally intensive**: The use of LOO validation in this study was possible because the largest number of features in any single subset was only 72 (Optimal Feature Subset A of the one year forward forecast model). Therefore, the small sample of companies coupled with the limited number of features made the use of LOO validation feasible with regards to the resources available.

- **Each independently trained classifier needs to be sufficiently similar**: This is the case with both the kNN and KRR techniques (as opposed to the traditional Multi-Layer Perceptron used in other Neural Network studies). This characteristic has been discussed in detail in Chapter Nine.

19.4. TESTING THE SEPARABILITY INDEX VALUE AS A MEASURE OF 1NN CLASSIFICATION ACCURACY

An additional test was performed in order to determine how closely the SI value matched the classification accuracy rate of the nearest neighbour classifier.

As described in Chapter Eighteen, ten different feature subsets were collected for each of the three prediction models. A random selection of fifteen of the thirty possible subsets was input into a 1NN classifier using leave-one-out (LOO) validation.

The accuracy rates achieved using the 1NN were identical to the SI values in all cases. As described before, this makes intuitive and mathematical sense. The SI values have been disclosed in Appendix H.

19.5. THE METHODOLOGY APPLIED IN THE CONSTRUCTION OF THE KNN CORPORATE FAILURE PREDICTION MODEL

The following steps were followed in applying the kNN methodology to the construction of each of the one, two and three year forward corporate failure prediction models in this study:

(i) Determination of the optimal value of k
(ii) Construction of the optimal kNN classifier
(iii) Evaluation of the optimal kNN classifier

19.5.1. DETERMINATION OF THE OPTIMAL VALUE FOR K

In order to determine the optimal value for k, the kNN classifier was run using all odd values of k between 1 and 21. Only odd values were used in order to avoid the situation in which a test point has an equal number of training points of both failed and non-failed classes as its nearest neighbours (i.e. a split vote).

The optimal feature subset(s) for each model were input into the classifier and the accuracy of each model was assessed using leave-one-out (LOO) validation.

The Matlab code used to perform this procedure has been included in Appendix A.3. The function "testXfoldknn" was used, setting variables *startk* to 1 and *endk* to 21. C was set equal to the number of input vectors sets included in the input matrix *X* in order to implement LOO validation.

In this way, the value of k that resulted in the greatest prediction accuracy was determined.

19.5.2. Construction of the Optimal kNN Classifier

Using the optimal value of k, the kNN classifier was run on each optimal feature subset for each of the one, two and three year forward prediction models.

This was performed using the Matlab code for the function "XfoldknnYT" included in Appendix A.3. The input variable, k, was set to the value determined as described above. C was set equal to the number of input vector sets included in the input matrix X in order to perform LOO validation.

19.5.3. Evaluation of the Optimal kNN Classifier

The accuracy of each model was determined using LOO validation.

The error rates were split between type I and type II errors. A type I error is one in which a company that has actually failed is predicted to be a non-failure. Conversely, a type II error is one in which a non-failed company is predicted to be a failure.

	Actual Failure	Actual Non-Failure
Predicted Failure	Correct	**Type II Error**
Predicted Non-Failure	**Type I Error**	Correct

Table 19.2. Illustration of the type I and type II errors that can be incurred.

The percentages of type I and type II errors are calculated as the probabilities of errors conditional on the actual status of the firm. For example:

Type I error rate = (No. of type I errors) / (No. of failed companies in sample)

These type I and type II error rates are calculated from the Y and T vectors output from the "XfoldknnYT" function (Appendix A.3.). The Y vector is the predicted class of the company as per the kNN classifier, while T is the true class.

The relative costs of these types of errors have been considered in Chapter 21.

19.6. RESULTS OF KNN MODEL CONSTRUCTION

19.6.1. RESULTS: OPTIMAL VALUE OF K

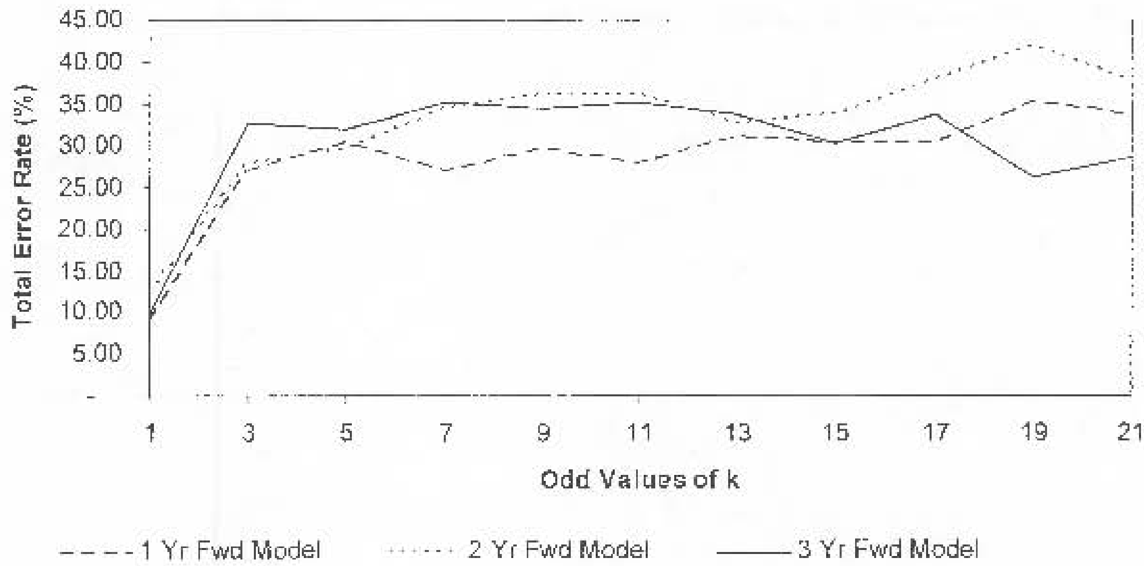


Figure 19.2. Total kNN error rates for the 1, 2 and 3 year forward prediction models over multiple values of k.

The above graph depicts the total error rates for all three prediction models. Optimal feature subsets for each forecast period performed identically to one another.

It is clear that the optimal value of k (i.e. the value of k that results in the lowest error rate) is k=1 for all three models. This is the basic nearest neighbour classifier (1NN).

	Optimal Value of *k* (*Total Error Rate*)		2nd Best Value of *k* (*Total Error Rate*)	
1-Yr Fwd Model	1	(9.02%)	3 & 7	(27.05%)
2-Yr Fwd Model	1	(12.30%)	3	(27.87%)
3-Yr Fwd Model	1	(9.84%)	19	(26.23%)

Table 19.3. The best and second best performing values of *k* using Optimal Feature Subset A for each prediction model (Total error rates in brackets).

The second best model in each case is more than three times as inaccurate as the optimal nearest neighbour model.

19.6.2. Results: Classification Accuracy of Optimal kNN Classifier

The type I and type II error rates incurred on the optimal nearest neighbour classifier constructed are as follows.

	Type I Error Rate	Type II Error Rate	Overall Error Rate
1-Yr Fwd Model			
- Optimal Feature Subset A	9.84%	8.20%	9.02%
2-Yr Fwd Model			
- Optimal Feature Subset A	6.56%	18.03%	12.30%
- Optimal Feature Subset B	11.48%	13.11%	12.30%
3-Yr Fwd Model			
- Optimal Feature Subset A	11.48%	8.20%	9.84%
- Optimal Feature Subset B	11.48%	8.20%	9.84%
- Optimal Feature Subset C	9.84%	9.84%	9.84%

Table 19.5. The type I and type II error rates incurred in the prediction of corporate failure using the optimal nearest neighbour (1NN) classifier.

The error rates are further evaluated in Chapter 21.

CHAPTER 20
CORPORATE FAILURE PREDICTION MODEL: KERNEL RIDGE REGRESSION

20.1. INTRODUCTION

20.1.1. THE LEARNING PROBLEM

> "The problem of understanding intelligence is said to be the greatest problem in science today and "the" problem for the century." (Poggio & Smale, 2003, 537).

Poggio & Smale argue further that learning represents the gateway to understanding this intelligence. The general notion of learning problems is to find a rule, which, based on external observations, assigns an observation to one of several classes (Muller, Ratsch, Tsuda & Scholkopf, 2001, 181).

Machine learning algorithms establish such "rules" by learning from examples (known as "inductive learning").

Through inductive learning, a machine can predict an unknown attribute or classification by generalising from a collection of fully-described objects. In most learning problems, a few hundred to a few thousand known cases are needed in order for a machine to accurately learn and perform such induction.

However, both people and animal's brains are able to learn from just a few examples. Discovering how the human brain works will allow for the construction of intelligent machines that learn from experience and improve their competencies as children do.

The history of the development of machine learning techniques from brain-theory is discussed in greater depth in Chapter Nine.

20.1.2. STATE-OF-THE-ART INDUCTION ALGORITHMS

Historically, many traditional neural networks have been applied to the task of corporate failure prediction. However, as explained in Chapter Nine, there are a number of fundamental drawbacks to such an approach:

- it leads to difficult optimisation,

- there are uncertainties relating to network structure, and
- often it can not identify the global optimum, inhibiting the use of cross-validation techniques.

Kernel methods, with radial basis kernels placed on each data point, can be viewed as a form of rational reconstruction of the neural network concept. The kernel approach:

- simplifies the training process,
- is, in many respects, a self-designing network, and
- will always lead to one unique and consistently identified optimum.

However, the origins of kernel methods can also be derived from earlier statistical modelling processes. This is the angle from which they will be approached in this chapter.

Support vector machines (SVMs), a kernel method, are currently enjoying a wave of popularity and are considered by many to be state-of-the-art classifiers (Cristianini & Shawe-Taylor, 2000). They are based on statistical learning theory (Vapnik, 1998) and exhibit extremely good empirical performance (Leich & Hornik, 2003).

The wider application of SVMs is inhibited by their seemingly complex nature – explanations of their functioning in the literature require the understanding of such topics as infinite-dimensional Hilbert space and functional analysis. Cristianini & Shawe-Taylor (2000) provide a rigorous standard explanation of this nature.

However, in an interesting recent development it has been shown by Rifkin (2002) and Poggio & Smale (2003) that SVMs are much more closely related to classical statistical methods, such as ridge regression, than has hitherto been believed. Moreover, and rather surprisingly, it turns out that such simpler classical methods, applied with care and understanding, are capable of performance that is competitive with, and in certain situations better than, support vector machines.

20.1.3. Structure of this Chapter

This chapter starts with a simple and intuitive introduction to classification using kernel ridge regression (KRR). This explanation is decomposed into a discussion of:

- ordinary regression,
- kernel-based regression, and finally
- ridge regression.

Thereafter, the theoretical section of this chapter proceeds on to show how KRR resembles and differs from SVMs.

Finally, the chapter concludes with the presentation of the empirical results obtained in the application of KRR to corporate failure prediction in this study. This discussion addresses the determination of the parameters for the optimal KRR model and tabulates the prediction results achieved using this model.

20.1.4. Notational Conventions

The description of KRR and its implementation is algebraically rigorous. The following table presents the notational conventions used in the ensuing discussion.

Algebraic Notation	Explanation
P	Number of data points (i.e. the number of cases).
N	Number of features in each input vector x_i.
x_i^a	Feature a of input vector x_i of data point / case P_i.
y_i / t_i	The target / class label of P_i.
n and p	The number of +1 and -1 points in a data set, respectively.
$L(f(x),t)$	An empirical loss function.
$K(x,x') = K(x_i,x)$	A local basis function localised at x_i.
$\|x_i - x_j\|$	The Euclidean distance between x_i and x_j.
σ	Sigma, the kernel width (RBFs).
α	Alpha, the kernel weight.
$\|f\|^2$	The norm of the input vector.
γ	Gamma, the regularisation constant.
$\mathbf{a}, \mathbf{K}, \boldsymbol{\gamma}, etc$	Matrix notation: the alpha, kernel and gamma matrix, etc.
$\mathbf{I}$	Identity matrix.

Table 20.1. Notational conventions for discussion relating to KRR

20.2. DISCUSSION OF KERNEL RIDGE REGRESSION (KRR)

20.2.1. ORDINARY REGRESSION

(a) Defining classification by regression

Regression is a process of fitting a curve to a set of data points with a view to predicting a new point. Given a set of P data points $(x_i, y_i)_{i=1}^{P}$, regression is the process that seeks to estimate a smooth function $y = f(x_i)$ consistent with these points. New results are merely interpolated from the data supplied.

The response variable can be a continuous variable, y , (in the case of regression or modelling) or it can be a discrete label, t, (e.g. ±1) as in a classification task such as performed in this study:

- In **modelling**, when given a new point x_{new}, $f(x_{new})$ should directly estimate the y-value associated with it.
- In **classification**, the sign of the function $t_{new} = sign(f(x_{new}))$ would serve as the prediction.

There is no essential difference between the processes of classification and modelling, except that in classification:

- the y-values of the training data are restricted to discrete values, and
- the label assigned to a new value, x_{new}, is not $f(x_{new})$ but $sign(f(x_{new}))$

Illustrative classification example

In reality, the vector x_i will probably be described by N numerical values (i.e. an N-dimensional vector or $\Re^N$). However, there is no loss of generality in describing the case of $\Re^1$. As explanations in lower dimensions are easier to understand and diagram, most of the explanations in this chapter will be represented by examples in these lower dimensions. For $N > 1$ nothing changes except that $f(x)$ becomes a function of many variables rather than a single one. The extension is handled automatically by using vector/matrix notation and a matrix-oriented computer language such as Matlab.

Hence, continuing with the example of the classification of men and women used in Chapter Seventeen, consider the simplified case in which this classification can be performed based on the single attribute of waist-to-hip ratio (x). Plotting x along a single axis (with women as circles and men as crosses) may appear something like this:

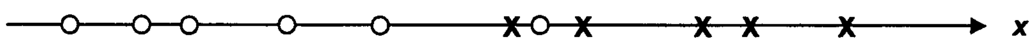


Figure 20.1. An arbitrary plot of men (crosses) and women (circles) on a single x axis.

The objective of the exercise is to create an induction algorithm that is a smooth function of x and which has a positive value in the vicinity of crosses and a negative value in the vicinity of circle points (assuming that the dichotomous male-female classification is coded as such). The function would cross over from negative to positive at some threshold value:

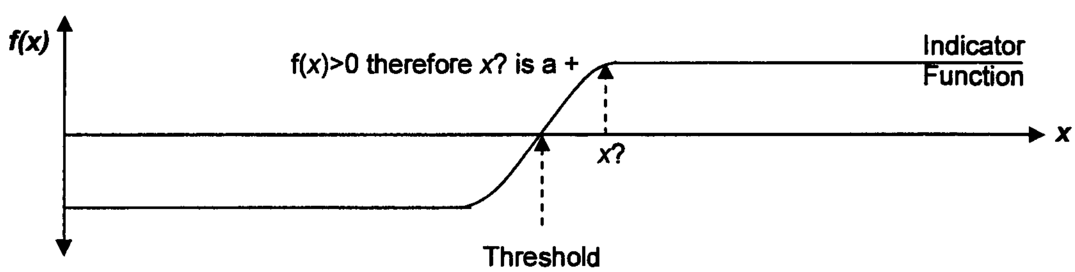


Figure 20.2. An example of a suitable indicator function for dichotomous classification based on a single attribute, x.

(b) Curve fitting

The following steps define the general strategy for fitting a curve to a set of training data points:

- **Step 1: Flexible "universal" functions**
  A flexible "universal" function, with adjustable parameters, needs to be selected.

  The form of the function is, in principle, arbitrary as long as it is a "universal" function in that it can take on any required shape by a suitable adjustment of its parameters (Greene, 2003). In practice, some functional forms are more suitable than others (for example, a Fourier function works well if the data represents a periodic function; otherwise it is less satisfactory). A simple and highly effective strategy is to use a radial basis function (RBF) (discussed further below).

- **Step 2: Empirical loss functions**

  The empirical loss function, $L(f(x),t)$, is some function of the discrepancy between the training target values predicted by the function and the actual target values associated with the training points (Bishop, 1995, 9). Using some kind of optimisation algorithm, the parameters can be adjusted so that this "measure of mismatch" is minimised.

  The optimisation algorithm that is used to minimise the loss function in order to fit the universal function to the training points, will depend on the form of the applied loss function.

  There are many possible loss functions that may be applied when performing regression. The common least squares loss function is the square distances from the function to the training points summed over all the training points:

$$L(f(x),t) = \sum_{i=1}^{P}(f(x_i) - t_i)^2 \tag{20.1}$$

  Squaring the deviations ensures that all errors are non-negative and, therefore, cannot cancel one-another out. In addition, this loss function gives increased weight to bigger deviations.

  The hinge loss function, applicable in classification problems, is a simple count of the number of misclassifications ((Cristianini & Shawe-Taylor, 2000, 77):

$$L(f(x),t) = \sum_{i=1}^{P}(sign(f(x_i) - t_i)) \tag{20.2}$$

20.2.2. Kernel-Based Regression

As mentioned above, the universal function can take many forms. However, a simple and effective one is a function which is a weighted sum of local basis functions located at each training point.

(a) The kernel

A local basis function or "kernel" is simply a function of the distance between two points (Greene, 2001b). A kernel can take many forms:

Kernel name	Kernel Function
Gaussian RBF	$K(x,x') = e^{-\lVert x-x' \rVert^2 / 2\sigma^2}$
Polynomial	$K(x,x') = ((x.x') + \vartheta)^d$
Sigmoidal	$K(x,x') = \tanh(\kappa(x.x') + \vartheta)$
Inverse Multiquadratic	$K(x,x') = \dfrac{1}{\sqrt{\lVert x-x' \rVert^2 + c^2}}$

Table 20.2. Common kernel functions: Gaussian RBF, polynomial, sigmoidal and inverse quadratic kernel functions (where *c*>0). (Source: Muller et al, 2001, 184)

Specifically, this study seeks to employ a local basis function $y = K_i(x) = K(x_i, x)$ located at x_i, that returns a value of unity at $x_i = x$ and tends smoothly to zero as x departs from x_i. Using conventional notation, these properties can be expressed as follows:

- When $x_i = x$, $K(x_i, x) = 1$.
- For large $\lVert x_i - x \rVert$, $K(x_i, x) \to 0$. (In order to allow for a multidimensional x, the Euclidean distance $\lVert x_i - x \rVert = \sqrt{\sum (x_i - x)^2}$ is used instead of the absolute difference $|x_i - x|$.)

A kernel that is symmetric in this way (i.e. its rate-of-fall-off is the same in all directions) is known as a radial basis function (RBF) (Greene, 2001). The Gaussian kernel is an example of this:

$$K(x_i, x) = e^{-\frac{\lVert x_i - x \rVert^2}{2\sigma^2}} \quad (20.3.)$$

The Gaussian kernel is a kernel that occurs almost universally in the literature. Bishop (1995, 164) dealt with the application of Gaussian kernels across various fields of study. It makes an intuitively good choice because, as the distance from the centre increases, the value of the function decreases smoothly and tends to zero for large distances. Further motivation for the use of radial basis functions for function approximation, comes from the theory of kernel regression (Scott, 1992). This theory shows that the use of RBFs for estimating regression functions from noisy data can be motivated based on methods of kernel density estimation (see Figure 20.3. below).

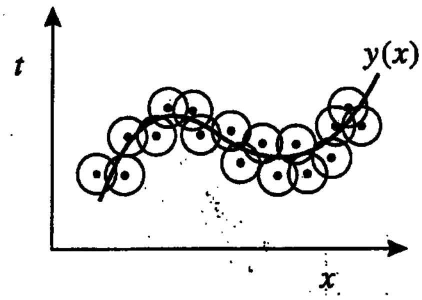


Figure 20.3. Schematic illustration of the use of a kernel estimator to model the joint probability density in the input-output space.

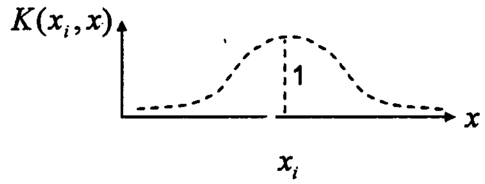
The dots show the data points, and the circles represent Gaussian kernel functions centred on the data points, while the curve shows the regression function given by the conditional average of *t* as a function of *x*. (Source Bishop, 1995, 178)

(b) Construction of a kernel-based universal function

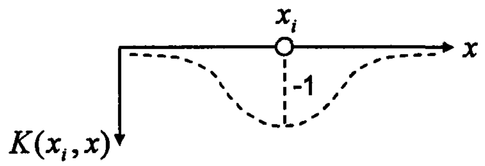
The function $f(x)$ is constructed as a weighted sum of these basis functions.

Continuing with the example of the classification of men and women started above (Figure 20.2.), this can be illustrated graphically as follows:

- A kernel ($K(x_i,x)$) is centred on each observation x_i in the training set, so that its value is +1 where $x = x_i$ and falls off smoothly towards zero as x deviates from x_i. Note that the constant, σ, determines the width of such "bump".



- Each kernel is then multiplied by its classification label t_i. This has the effect of inverting the kernels associated with the class which has been denoted as negative (women in this example).



- All the basis functions associated with each training observation are then added together and weighted in order to create the indicator function.

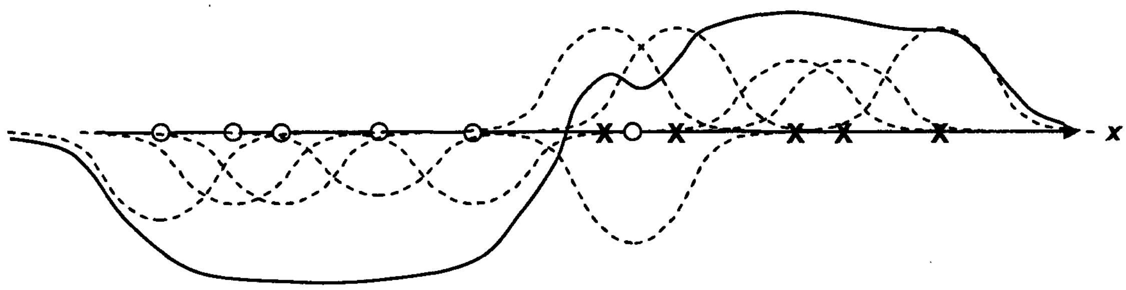


Figure 20.4. Illustration of the construction of a weighted kernel indicator function.

Negative kernels are associated with women (circles) and positive kernels with men (crosses). The classification of an unknown observation would be the sign of the indicator function at that point.

Algebraically, the $f(x)$ function illustrated above can be represented as:

$$f(x) = \sum_{i=1}^{P} \alpha_i . t_i . K(x_i, x) \qquad (20.4.)$$

In order to get the best results, the training data should be balanced (i.e. there should approximately be an equal number of +1 and –1 observations). If this is not the case, the boundary of the induction algorithm will be shifted further away from the class with the greater number of observations. This is as a result of the greater value of the sum of the kernels associated with the larger group.

Balance can be imposed by modifying +1 and –1 to $(n+p)/n$ and $-(n+p)/p$ respectively (where n and p represent the number of –1 and +1 points, respectively) (Duda & Hart, 1973). However, no such bias existed in this study as a paired sample technique was employed.

(c) *Calculating the parameters of the kernel regression function*

There are two parameters that need to be calculated at this stage, namely α (the kernel weight) and σ (the kernel width).

The values of the α multipliers (coefficients) can simply be calculated by substituting the data points into the general equation, yielding a set of equations (one for each data point):

$$\begin{aligned}
&\alpha_1 t_1.k(x_1,x_1)+\alpha_2 t_2.k(x_1,x_2)+\ldots+\alpha_P t_P.k(x_1,x_P)=f(x_1)\\
&\alpha_1 t_1.k(x_2,x_1)+\alpha_2 t_2.k(x_2,x_2)+\ldots+\alpha_P t_P.k(x_2,x_P)=f(x_2)\\
&\alpha_1 t_1.k(x_3,x_1)+\alpha_2 t_2.k(x_3,x_2)+\ldots+\alpha_P t_P.k(x_3,x_P)=f(x_3)\\
&\ldots\\
&\alpha_1 t_1.k(x_P,x_1)+\alpha_2 t_2.k(x_P,x_2)+\ldots+\alpha_P t_P.k(x_P,x_P)=f(x_P)
\end{aligned} \tag{20.5}$$

In matrix notation, the set of equations in (20.5) can be written as follows:

$$K\times\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \ldots\\ \alpha_P\end{bmatrix}=\begin{bmatrix}f(x_1)\\ f(x_2)\\ f(x_3)\\ \ldots\\ f(x_P)\end{bmatrix} \text{ or simply } \mathbf{K.\alpha}=\mathbf{f(x)} \tag{20.6}$$

Since there is a basis function for each data point, the number of unknown parameters (α) is the same as the number of kernels (K). Thus, the system of equations is square and potentially solvable.

$$\boldsymbol{\alpha}=\mathbf{K}^{-1}.\mathbf{f(x)} \tag{20.7}$$

If σ is small and the kernels are narrow relative to the spacing between the data points, then solving for (20.7) will result in a curve that will pass exactly through each data point (interpolation). The indicator function will then behave in the same manner as a nearest neighbour classifier. This can be illustrated graphically by narrowing the kernels from Figure 20.4. as follows:

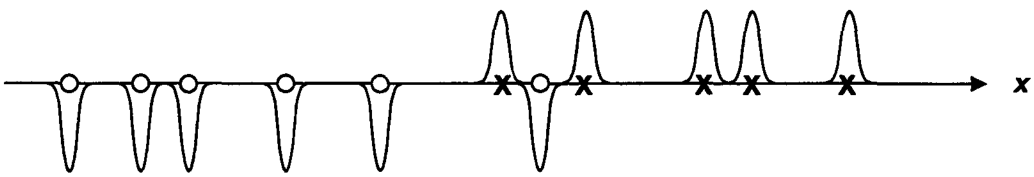


Figure 20.5. Illustration of a kernel indicator function with σ small relative to the spacing between the data points.

If, as in real life problems, the data has a degree of noise associated with it, the resultant curve will be highly irregular, as is illustrated above. The indicator function will fail to generalise on new data as it will fit the particular noise of the training data set rather than mapping the stable, causal relationships embedded in it. This is known as over-fitting and is discussed in more detail in Chapter Nine.

Bishop (1995, 166) showed that "smoothness" can be enforced on the fitted curve by increasing the width of the kernels. He also showed that if the value of σ is set very large, the model will approach that of a simple linear classifier (see below). If the data points are not linearly separable, then such a classifier will also generalise poorly.

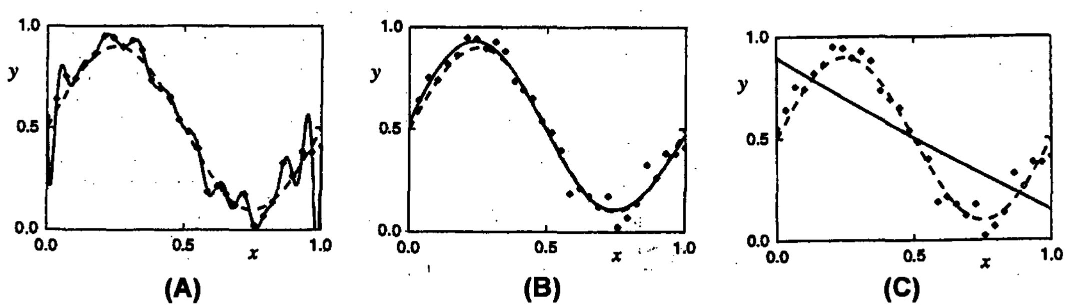


Figure 20.6. A simple example of how σ impacts on the fit of the curve to a set of data points when using radial basis functions.

In the example in Figure 20.6., a set of 30 data points was generated by sampling the function

$$y = 0.5 + 0.4\sin(2\pi x),$$

shown by the dashed curve in the graphs above, and adding Gaussian noise with standard deviation of 0.05. The solid line shows the fitted curve in each case. (Source: Bishop, 1995, 166-184)

- **(A):** A simple example of exact interpolation using radial basis functions. The width parameter σ was set to 0.067 (corresponding to roughly twice the distance between data points).
- **(B):** A curve that generalises better than (A) based on the original sampling function. The width parameters of the basis functions were set to σ =0.4.
- **(C):** The width parameter was set large to σ =10. The function is over-smoothed resulting in a poor representation of the underlying function that generated the data.

20.2.3. Ridge Regression

(a) The need for regularisation

Although smoothness can be enforced on a curve by increasing the value of σ, as shown above, the equation solving process will still try to force an exact fit, leading to an ill-conditioning of the solution-process and numerical instability.

Regression is an "ill-posed" task: there are an infinite number of continuous lines that can pass through any finite number of points. On the one extreme, a straight line can always be fitted to the data using the "least-squares" loss function. However, if the data is not linearly separable, this line will not generalise well on unseen data points. At the other extreme, a line which passes exactly through all the points will overfit the data, fitting the random errors and idiosyncrasies of the particular training data sample. Inbetween these two extremes are an infinite number of lines which compromise between the competing ideals of:

- minimising the error of the fit and
- smoothness.

Focusing exclusively on minimising the loss function described in equation (20.1.) only addresses the "error of fit" ideal. When using a flexible universal function, this might lead to a highly complex function $f(x)$ which happens to fit the training points closely but which has little predictive value for points outside of the training data.

Regularisation is a way of controlling the "smoothness" property of a mapping function (Bishop, 1995, 171). It involves adding to the loss function an extra term which is designed to penalise mappings proportionally to their complexity (Greene, 2001). An additional regularisation term added to the empirical loss function should typically give rise to larger values when $f(x)$ is more complex.

One way of doing this is by "Tikhonov regularisation" (Tikhonov & Arsenin, 1977), which was introduced into learning theory by Poggio & Girosi (1990). Using this method, smoothness is enforced on $f(x)$ by simultaneously minimising $L(f(x),t)$ and its norm, $\|f\|^2$.

$$E = L(f(x),t) + \gamma\|f\|^2 \tag{20.8.}$$

The norm, which measures the length of the input vector, is larger for more complex functions (i.e. functions with a greater $\Re^N$), resulting in a greater penalty term being added to the error function.

The Representer Theorem (Girosi, 1998; Schölkopf et al, 2001) proves that there is a solution to the general equation (20.8) that has the form of equation (20.4.)

The regularisation constant, γ, controls the relative importance of the penalty term and, therefore, the degree of smoothness imposed on the function $f(x)$. This constant, therefore, manages the trade-off between the fit and smoothness (or bias and variance, as it is termed in statistics) (Poggio & Smale, 2003, 541).

In empirical work, the optimum γ value can be found using cross-validation on a trial-and-error basis (Wahba, 1990).

(b) Algebraic explanation of ridge regression

A solution to the ill-conditioning of the regression solution process and numerical instability has long been known in statistics and numerical mathematics: increase the positive magnitude of the leading diagonal of the matrix $\mathbf{K}$ by adding to it a small multiple of the identity matrix (Greene, 2003). This seeks to improve numerical stability and is also useful when $\mathbf{K}$ is not of full rank (Cristianini & Shawe-Taylor, 2000, 22).

Once again, the regularisation constant, γ, controls the trade-off between low loss squares (fit) and low norm (smoothness) in the solution (Cristianini & Shawe-Taylor, 2000, 22).

In statistics this method is called "ridge regression" or "shrinkage" since it tends to shrink poorly-determined α-values toward zero. Adjusting equation (20.7) for the ridge terms results in (Cristianini & Shawe-Taylor, 2000, 23):

$$\boldsymbol{\alpha} = (\mathbf{K} + \gamma\mathbf{I})^{-1}.\mathbf{f}(\mathbf{x}) \tag{20.9}$$

Cristianini & Shawe-Taylor (2000, 119) showed that equation (20.9.) can be derived from the Tikhonov regularisation square loss function described by equation (20.8).

(c) Intuitive explanation of ridge regression

Ridge regression is well known as a solution to the problem of multicollinearity in multivariate linear regression.

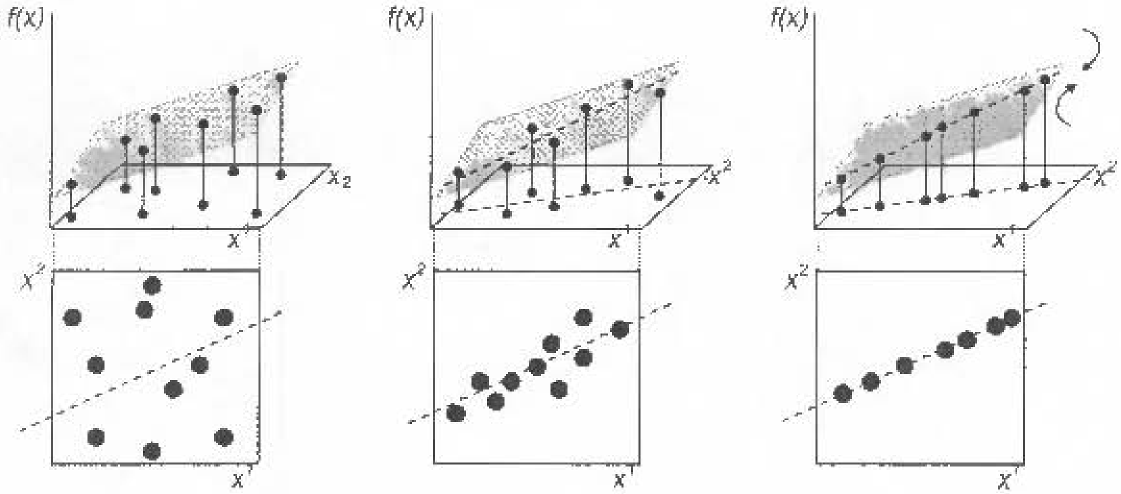
This can be illustrated with the problem of fitting a plane:

$$f(x_i) = \alpha_0 + \alpha_1 x_i^1 + \alpha_2 x_i^2 \qquad (20.10.)$$

to a set of points $X = \{x_i^1, x_i^2\}$ (see Figure 20.7. below).

If x_i^1 and x_i^2 are not correlated at all, the points will be scattered over the input plane, $\mathbf{X}$, and the fitted plane in $\Re^3$ (i.e. both α_1 and α_2) will be well-defined (Diagram (A) below).

If, however, the features x_i^1 and x_i^2 are correlated, then the distribution of points on $\mathbf{X}$ will fall in a narrow ellipse. As a result, the slope of the fitted plane in the direction perpendicular to the major axis of this ellipse (dotted line in Figure 20.7.) will be poorly-defined (Diagram (B) below). Taking this situation to its extreme, in the case of perfectly correlated features, there would be an infinite number of planes and a unique solution would be impossible (Diagram (C) below).



(A) Input data uncorrelated:
Due to the "spread" the tilt of the fitted plane is well-defined in all directions.

(B) Input data partially correlated:
The fitted plane is rather unstable about the vertical axis shown.

(C) Input data highly correlated:
The fitted plane is highly unstable about the vertical axis, as indicated.

Figure 20.7. Graphical representation of the impact of a poorly-conditioned input matrix, *X*, on fitting a solution plane. (Source: Adapted from Greene, 2002)

The input matrix $\mathbf{X}$ in (A) above implies that the columns of $\mathbf{X}$ are independent of each other. The matrix is said to be of "full rank". In the correlated case, (C), the columns of $\mathbf{X}$ are linearly dependent on each other and the matrix is said to be of less than full rank. In (B), $\mathbf{X}$ is "poorly-conditioned" resulting in an "exaggeration of rounding errors and poor solution accuracy" (Greene, 2002).

In Chapter Seventeen, several methods for addressing multicollinearity are discussed. These include:

- Identifying redundant inputs in order to eliminate them;
- Using vector algebra, such as principle component analysis and singular value decomposition, to project the data onto a smaller set of mutually orthogonal axes;
- Using a method called ridge regression, as employed in this study.

The ridge regression method imposes a bias on the fitted hyperplane in the direction in which the separator is poorly defined. This can be achieved by appending to the centred training data a set of virtual points close to the origin, one on each axis, for example [{k,0,0,0…,0}, {0,k,0,0,…,0}…{0,0,0,0,…,k}].

This has the effect of reducing the slope of the hyperplane in all directions. It will, however, have a negligible impact on the slope in the directions that are well defined and increasing impact on those axes in which the solution is unstable.

20.2.4. Kernel Ridge Regression

Equation (20.9.) represents the algorithm for kernel ridge regression. Cristianini & Shawe-Taylor (2000, 119) noted that this algorithm has appeared independently under a number of different names. It is also known as Kriging and the regularised least squares classifier (RLSC).

20.3. THE RELATIONSHIP BETWEEN KRR AND SVM

20.3.1. Re-emergence of an Old Idea

KRR is capable of extremely good performance in real-world classification tasks. There have been many recent studies (Poggio & Smale, 2003; Rifkin, 2002; Fung & Mangasarian, 2001; Suykens, Van Gestel, De Brabanter, De Moor & Vandewalle, 2002) which show that its performance closely follows that of the state-of-the-art, almost exactly emulating that of better-known contenders such as the SVM.

Number of documents per class used for training:	SVM	RLSC
800	0.131	0.129
250	0.167	0.165
100	0.214	0.211

Table 20.3. A comparison of SVM and RLSC accuracy on a multi-class classification task (the *20newsgroups* dataset with 20 classes and high dimensionality, around 50 000), performed using the standard "one versus all" scheme based on the use of binary classifiers. Entries in the table are the fraction of misclassified documents. (Source: Rifkin, 2002)

Number of documents per class used for training:	SVM	RLSC
52	0.072	0.066
20	0.176	0.169
10	0.341	0.335

Table 20.4. A comparison of SVM and RLSC accuracy on another multi-class classification task (the *sector105* dataset, consisting of 105 classes with dimensionality about 50 000). Entries in the table are the fraction of misclassified documents.

In fact, KRR (or RLSC) owes its current renewed popularity to the SVM. KRR has recently been rediscovered through a series of modifications and developments within the SVM paradigm (for example Fung & Mangasarian's "Proximal SVM" (2001) and Suykens' "Least-squares SVM" (2002)). The treatment of the topic by these authors has tended to obscure the essential similarity between these new forms of the SVM and the much older idea of RLSC. This has been clarified by Rifkin in his MIT doctoral thesis ("Everything Old is New Again", 2002) and Poggio & Smale's paper ("The Mathematics of Learning") (Greene, 2003).

20.3.2. Connection Between KRR and SVM

The SVM uses the hinge loss function (equation (20.1.)) instead of the square loss function (equation (20.2.)). The resultant non-linearity means that the α -values cannot be determined by linear algebra. Instead, a quadratic optimisation program needs to be used.

The simplest implementation of this is the kernel Adatron (FrieB and Harrison, 1999). Here, the α -values are iteratively adjusted, one at a time, in such a way as to force the induction function to approximate the data points. Once again, the danger of over-fitting exists. The weights of the kernels associated with rogue outliers may become so large in the iterative process that the function becomes overly complex and over-fits the training points. This can be illustrated by increasing the weight of the kernel of the outlier from the example in Figure 20.4. above, as follows:

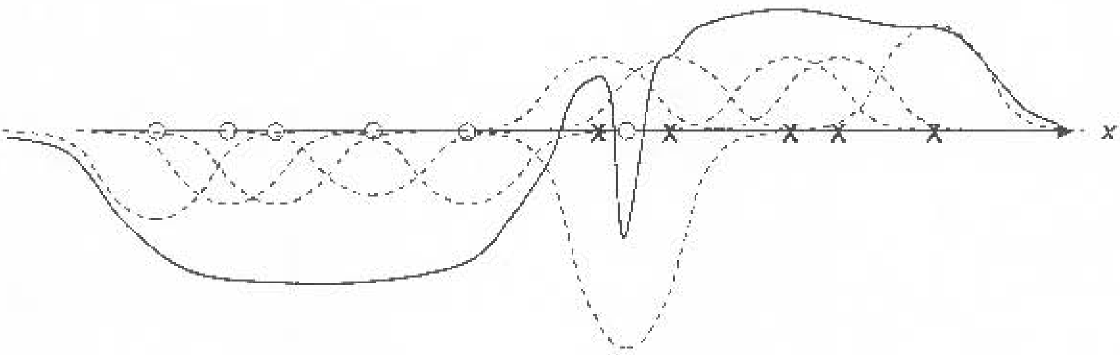


Figure 20.8. The impact of unregularised quadratic optimisation in the case of a hinge loss function SVM.

Function complexity can be controlled by placing an upper bound, C, on the weighting associated with any specific kernel (α). It can be shown that this is equivalent to applying ridge regression in KRR when adding $\gamma\mathbf{I}$ to the matrix $\mathbf{K}$. In fact C and γ are related as follows (Rifkin, Yeo & Poggio, 2003, 141; Cristianini & Shawe-Taylor, 2000, 84):

$$\gamma = \frac{1}{2}C \tag{20.11.}$$

20.3.3. Practical Differences Between KRR and SVM

If the regularisation parameters are correctly determined, there is little to choose between the two approaches in terms of generalisation accuracy (see Tables 20.3. and 20.4. above).

However, the use of the hinge loss function in the SVM tends to result in a more sparse solution, with many of the α-values becoming zero. Such a sparse model results in faster run-time classification and a lower data storage requirement.

Also, the iterative nature of the quadratic optimisation algorithm lends itself to partitioning schemes that allow the handling of vast quantities of data. In contrast, KRR, in its basic form, requires the storage and inversion of a P by P matrix. However, for moderately sized data sets, the simple KRR approach seems to be the preferred option (Greene, 2003).

20.4. THE METHODOLOGY APPLIED IN THE CONSTRUCTION OF THE KRR CORPORATE FAILURE PREDICTION MODEL

The following steps were followed in applying the KRR methodology in constructing each of the one, two and three year forward corporate failure prediction models in this study:

(i) Determination of the optimal values of the regularisation constant, γ, and the kernel width, σ;
(ii) Construction of the optimal KRR classifier;
(iii) Evaluation of the optimal KRR classifier.

The code for the KRR classifier is included in Appendix A.4. (the function "krr"). The notation next to the code explains how the KRR algorithm was applied. Note that the first step was to apply a Gaussian kernel to each training point before adding the ridge terms to the leading diagonal of the kernel matrix. Matlab was then used to solve for the regression weights.

20.4.1. DETERMINATION OF THE OPTIMAL VALUES OF THE REGULARISATION CONSTANT, γ, AND THE KERNEL WIDTH, σ

The regularisation constant, γ, and the kernel width, σ, are the two parameters that need to be user-selected in the construction of a KRR algorithm.

In order to determine the optimal values for γ and σ, the KRR classifier was run across a range of values for each parameter. On each run, the optimal feature subsets for each model were input into the classifier and the accuracy of each model, based on the given set of parameters, was assessed using leave-one-out (LOO) validation.

The following steps were followed in order to determine these optimal values on a trial-and-error basis:

- **Step 1**: The model was run across all values of σ from 0.05 to 2 in increments of 0.05. This range was deemed appropriate as all data was normalised prior to model construction (see Chapter Sixteen). γ was initially set to 1 in this step. σ was then plotted against the resultant KRR overall error rate in order to determine the values of σ at which the error rate was minimised. The code for this procedure is included in Appendix A.4. as function "testsigmaXfoldkrr".

- **Step 2**: A preliminarily determined optimal value of σ was then used to run the model across a wide range of exponentially increasing γ values:

  $$[10^{-6}, 10^{-5},\ldots, 10^{5},10^{6}]$$

  This was performed using the Matlab function "testgammaexp" included in Appendix A.4. The range of values for γ in which the overall KRR error rate was minimised, was then evaluated over smaller intervals in order to determine the optimum value for γ using the function "testgammaXfoldkrr".

  The accuracy of the models was not sensitive to changes in the value of γ (as presented and explained in the results tabulated below). As a result, a γ of 1 (as used in step (1) above) was maintained for the final step (C) in determining the optimal value of σ.

- **Step 3**: As γ remained the same as in step (1), the optimal value of σ was determined by running the function "testsigmaXfoldkrr" over the range of σ values determined as optimal in step (1) in increments of 0.005.

20.4.2. Construction of the Optimal KRR Classifier

Using the optimal values of γ and σ, the KRR classifier was run on each optimal feature subset for each of the one, two and three year forward prediction models.

This was performed using the Matlab code for the function "XfoldkrrYT" included in Appendix A.4. The input variables γ and σ were set to the values determined as described above. The input variable "*C*" in this Matlab function was set equal to the number of input vector sets included in the input matrix "*X*" in order to perform LOO validation.

20.4.3. Evaluation of the Optimal KRR Classifier

The accuracy of each model was determined using LOO validation. The error rates were split between type I and type II errors.

20.5. RESULTS OF KRR MODEL CONSTRUCTION

20.5.1. RESULTS: OPTIMAL VALUES OF γ AND σ

A wide range of values for γ were tested over a number of values for σ, all yielding no influence on the overall error rate of the models. γ is a regularisation parameter whose role is twofold (Greene, 2004):

- **To prevent outliers from locally distorting the decision boundary**: It plays an important role where the data are, in general, highly separable but where a small number of errant points or outliers deflect the decision boundary. In such a case one would expect to find a clear optimum for γ.

  If, on the other hand, there is an absence of well-defined outliers and the class clusters happen to form overlapping clouds, the role of γ is less clear and so it is likely have a clear optimum value. In such a case, the global smoothness of the decision boundary is more important than local deflections. Global smoothness is determined by σ, which has a clear optimal value in the case of this study.

- **To assist in removing numerical ill-conditioning**: This is described in the explanation of ridge regression above.

Therefore, the insignificance of γ to the construction of this classifier has the following possible implications:

- The available margin of the data set is dominated by class overlap rather than by the presence of identifiable outliers.
- The data set is numerically well-conditioned.

For this reason, γ was set equal to 1 for all KRR models constructed in this study.

The graphs below present the overall error rates for each of the KRR corporate failure models, setting $\gamma = 1$ and running each model over all values of σ from 0.05 to 2 in increments of 0.05 (as described in the previous section).

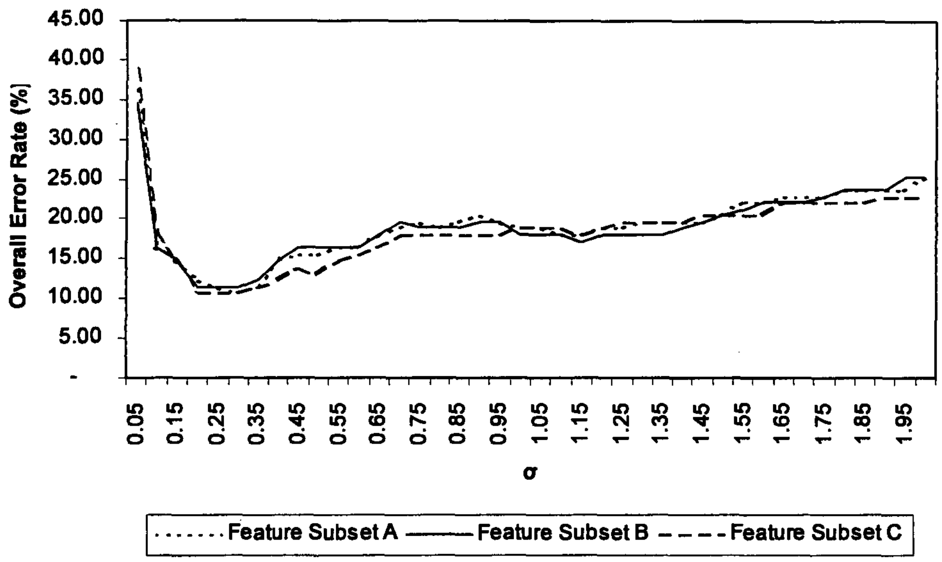


Figure 20.9. Graph of Three-Year Forward Prediction Model: Optimal σ search ($\gamma = 1$) for all feature subsets

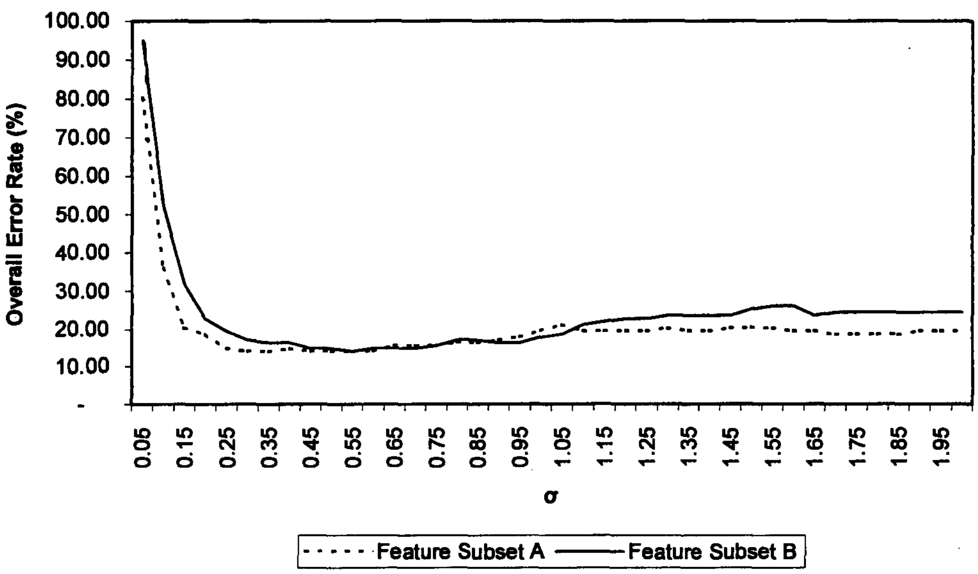


Figure 20.10. Graph of Two-Year Forward Prediction Model: Optimal σ search ($\gamma = 1$) for all feature subsets

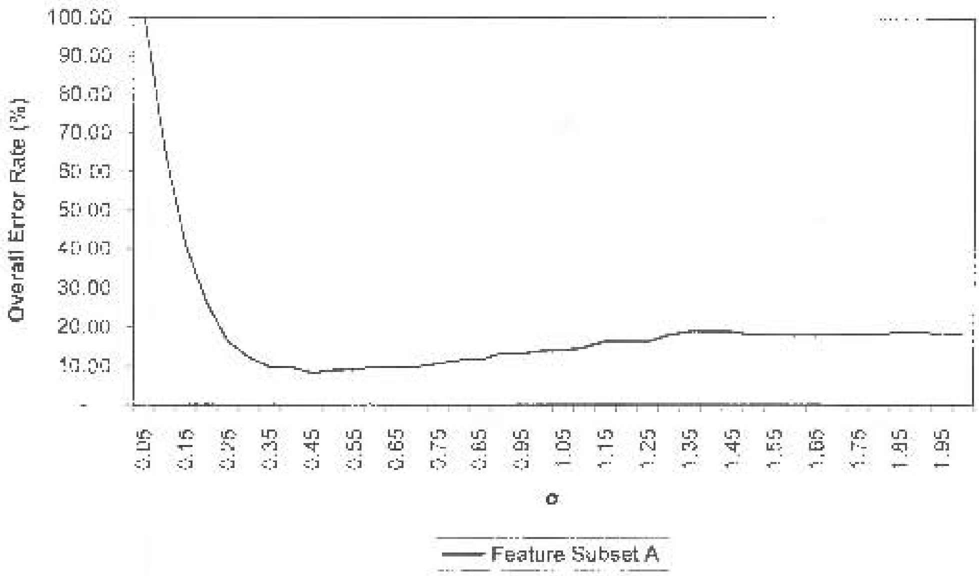


Figure 20.11. Graph of One-Year Forward Prediction Model: Optimal σ search (γ 1) for all feature subsets

It is clear from each of the graphs that there is a definite optimal range for σ where the error rate reaches a minimum. In order to determine the optimal σ value more exactly, the same process was repeated over the relevant minimum ranges of σ (as identified from the above tests) in increments of 0.005. These ranges were as follows:

	Lower Bound of Kernel Width	Upper Bound of Kernel Width
1-Yr Fwd Model		
- Optimal Feature Subset A	0.40	0.50
2-Yr Fwd Model		
- Optimal Feature Subset A	0.25	0.65
- Optimal Feature Subset B	0.50	0.60
3-Yr Fwd Model		
- Optimal Feature Subset A	0.25	0.35
- Optimal Feature Subset B	0.15	0.35
- Optimal Feature Subset C	0.15	0.35

Table 20.5. The optimal ranges of σ identified in the preliminary assessment of this parameter for each model.

The results achieved are presented below.

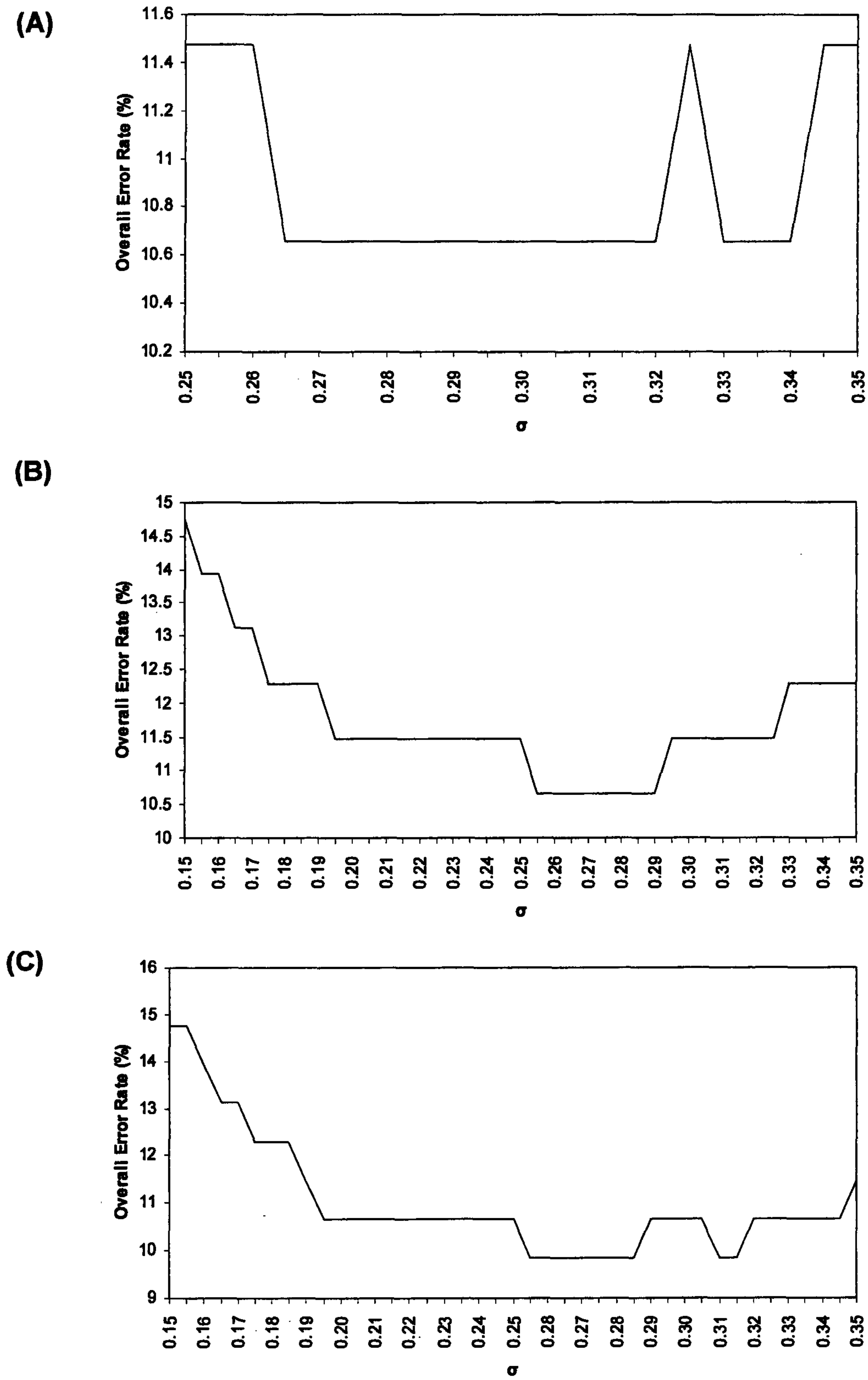


Figure 20.12. Graph of Three-Year Forward Prediction Models: Detailed σ search ($\gamma = 1$) (A: Optimal feature subset A; B: Optimal feature subset B; C: Optimal feature subset C)

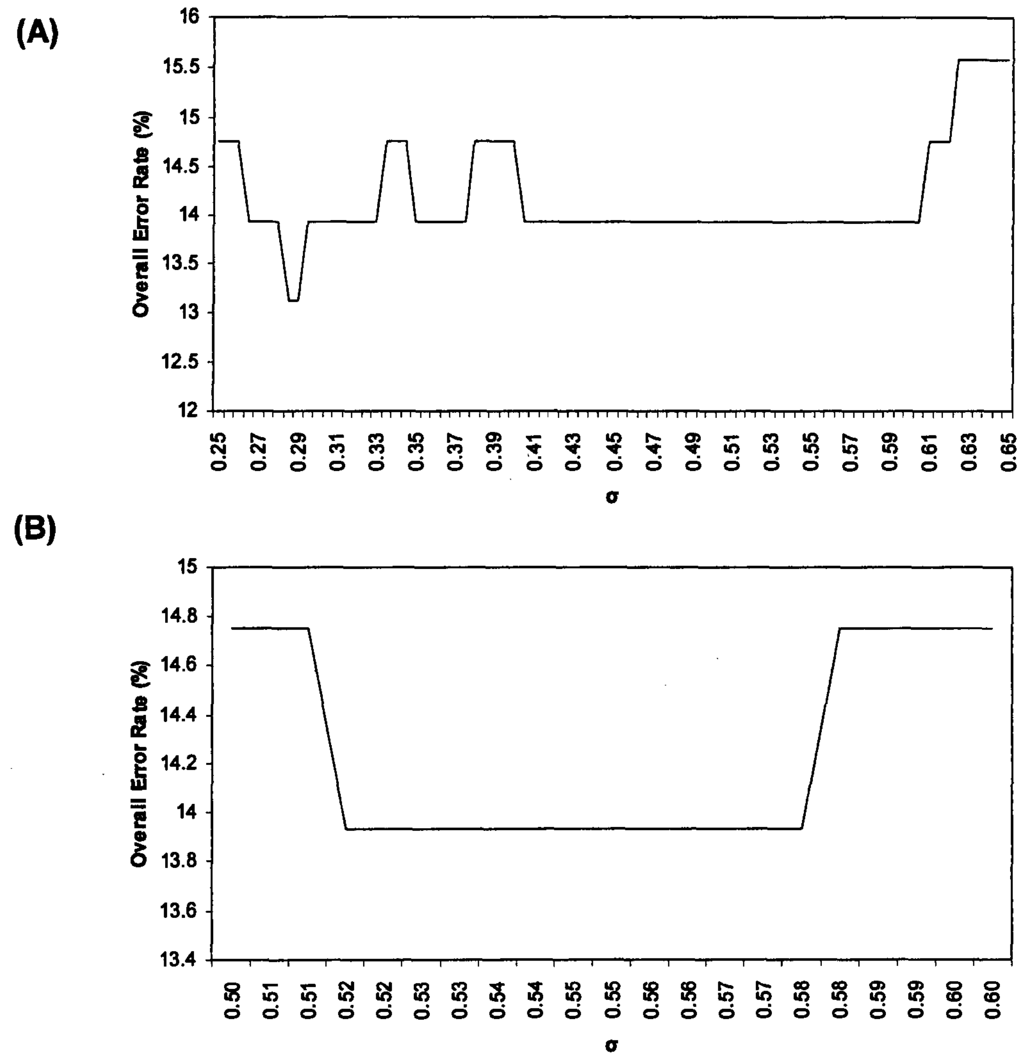


Figure 20.13. Graph of Two-Year Forward Prediction Models: Detailed σ search ($\gamma = 1$) (A: Optimal feature subset A; B: Optimal feature subset B)

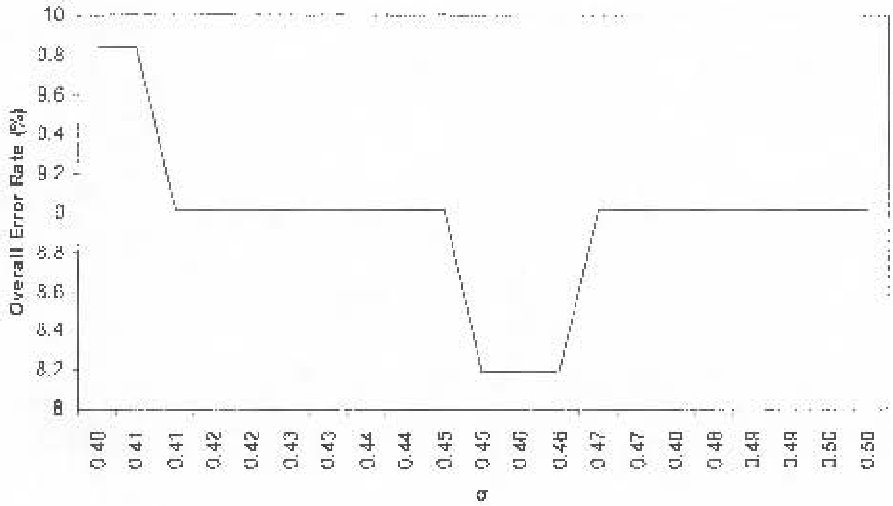


Figure 20.14. Graph of One-Year Forward Prediction Model (Optimal Feature Subset A): Detailed σ search ($\gamma = 1$)

The graphs above indicate clear ranges for σ in which the overall KRR error rate was minimised. The values that were selected for each parameter are presented in the following section.

20.5.2. Results: Classification Accuracy of Optimal KRR Classifier

The adjustable parameters and the type I and type II error rates incurred on the optimal KRR classifiers constructed were as follows.

	σ	γ	Type I Error Rate	Type II Error Rate	Overall Error Rate
1-Yr Fwd Model					
- Optimal Feature Subset A	0.45	1.00	8.20%	8.20%	8.20%
2-Yr Fwd Model					
- Optimal Feature Subset A	0.29	1.00	4.92%	18.03%	13.11%
- Optimal Feature Subset B	0.53	1.00	13.11%	14.75%	13.93%
3-Yr Fwd Model					
- Optimal Feature Subset A	0.27	1.00	11.48%	9.84%	10.66%
- Optimal Feature Subset B	0.27	1.00	11.48%	9.84%	10.66%
- Optimal Feature Subset C	0.27	1.00	9.84%	9.84%	9.84%

Table 20.6. The type I and type II error rates incurred in the prediction of corporate failure using the optimal KRR classifiers for each feature subset. The parameters that were selected in order to attain the tabulated results have also been included.

The error rates are further evaluated in the following chapter.

CHAPTER 21
EVALUATION OF THE CONSTRUCTED CORPORATE FAILURE PREDICTION MODELS

The preceding chapters have described the construction of two corporate failure classification models for each optimal feature subset for each forecast period – one model using k-Nearest Neighbours (kNN, Chapter Nineteen) and one model using Kernel Ridge Regression (KRR, Chapter Twenty).

This chapter describes the evaluation of these different models. Chapter Ten reviews a number of different methods for calculating error rates and evaluating model performance. However, some of the analysis described in Chapter Ten falls outside of the scope this study (discussed below).

The analysis described in this chapter incorporates type I and type II error rates, as well as a scenario analysis of which models performed best under different cost ratio assumptions. Finally, the misclassified companies are briefly examined for any trends applicable to model application.

Appendix I discloses the predicted class that each forecast model made for each company, along with the true class of that company.

21.1. LIMITATIONS OF EVALUATION

In the evaluation of the models that were constructed in this study, the following was noted:

- Inference of a percentage of total model errors to the population may not be meaningful in this study, as the probability of failure for the sample differs from that of the population.
- However, as a paired sample was used and, hence, both failed and non-failed companies each constituted half of the sample, it was not necessary to calculate a "weighted efficiency" measure (Korobow & Stuhr, 1984 – discussed in Chapter Ten).
- The research objective of this study was to construct a model for corporate failure prediction using a machine learning approach. Implicit in this research objective is the necessity to evaluate the different models constructed in order to determine

which model(s) are the best. However, an in-depth analysis of the actual model predictions in order to assess possible reasons for failure or misclassification does not fall within the scope of this report. Such an analysis was addressed insofar as it related to model comparison. Suggestions for further analysis are included in Chapter 23.

- The results attained by Altman (1968) in his seminal work on the construction of the Z-score model remain as a benchmark against which new studies compare their results. While other models have reported results that have outperformed those reported by Altman, these prediction accuracies have not been consistently, significantly or convincingly better (Zhang et al, 1999, 25). The performances of the models in this study are compared to the results attained by Altman. It should be noted, however, that direct comparisons to other studies can, potentially, be misleading because of the different time periods, geographic locations, assumptions and data availability under which different studies have constructed their prediction models (This has been discussed further in Section A).

21.2. SUMMARY OF RESULTS

The following table summarises the type I, type II and overall error rates for each model. These error rates have been summarised from Tables 19.5. and 20.6. for kNN and KRR, respectively.

Classifier	kNN			KRR		
Optimal Feat. Subset	A	B	C	A	B	C
1-Yr Fwd Model						
- Type I Err Rate	9.84%			8.20%		
- Type II Err Rate	8.20%			8.20%		
- Overall Err Rate	9.02%			8.20%		
- Misclass. Costs						
Cost Ratio 20:1	0.09758			0.08197		
Cost Ratio 1:1	0.09106			0.08197		
Cost Ratio 1:20	0.08275			0.08197		
2-Yr Fwd Model						
- Type I Err Rate	6.56%	11.48%		4.92%	13.11%	
- Type II Err Rate	18.03%	13.11%		18.03%	14.75%	
- Overall Err Rate	12.30%	12.30%		14.75%	13.93%	
- Misclass. Costs						
Cost Ratio 20:1	0.07104	0.11553		0.05543	0.13193	
Cost Ratio 1:1	0.12295	0.12295		0.11475	0.13934	
Cost Ratio 1:20	0.17486	0.13037		0.17408	0.14676	
3-Yr Fwd Model						
- Type I Err Rate	11.48%	11.48%	9.84%	11.48%	11.48%	9.84%
- Type II Err Rate	8.20%	8.20%	9.84%	9.84%	9.84%	9.84%
- Overall Err Rate	9.84%	9.84%	9.84%	10.60%	10.60%	9.84%
- Misclass. Costs						
Cost Ratio 20:1	0.11319	0.11319	0.09836	0.11397	0.11397	0.09836
Cost Ratio 1:1	0.09836	0.09836	0.09836	0.10656	0.10656	0.09836
Cost Ratio 1:20	0.08353	0.08353	0.09836	0.09914	0.09914	0.09836

Table 21.1. Summary of error rates and misclassification costs (under different cost ratio assumptions) for all constructed models

The calculation and analysis of the values in this table are discussed below.

21.3. RELATIVE ERROR RATES

21.3.1. MEASURING ERROR RATES

As discussed previously, Leave-One-Out (LOO) validation was used to calculate the error rates in this study. In other words, the classifier was trained on all data except a single case and then the constructed classifier was used to predict the class of that omitted case. This process was repeated (training the classifier omitting a different single case each time) until all cases had been predicted in this manner.

The overall error rate was calculated by dividing the number of misclassifications by the total sample size. Type I and type II error rates were measured as the probability of error conditional on the actual status of the firm (for example, the number of type I errors divided by the number of **actual** failures in the sample).

As discussed in Chapter Ten, a situation may arise in certain studies where the two sample groups (failed and non-failed) are significantly different in size. This can result in a distortion of the relative type I and type II costs, necessitating the use of a type of "weighted" efficiency measure to allow for meaningful comparison. As the sample in this study was split 50/50, it was not deemed necessary to calculate such a measure.

21.3.2. Analysis of Relative Error Rates

	Type I Error	Type II Error	Overall Error
1-Yr Fwd Model	KRR(A) kNN(A)	KRR(A), kNN(A)	KRR(A) kNN(A)
2-Yr Fwd Model	KRR(A) kNN(A) kNN(B) KRR(B)	kNN(B) KRR(B) kNN(A), KRR(A)	kNN(A&B) KRR(B) KRR(A)
3-Yr Fwd Model	kNN(C), KRR(C) kNN(A&B), KRR (A&B)	kNN(A&B) kNN(C), KRR(A,B&C)	kNN(A,B&C), KRR(C) KRR(A&B)

Table 21.2. Ranking of corporate failure prediction models by error rates (Optimal feature subset in brackets)

Table 21.2. ranks the various prediction models based on their error rates as presented in Table 21.1. Using the information in these two tables, the following was noted:

- **One Year Forward Model**:
  KRR outperformed kNN in terms of type I and overall error rate. The two methods were equal when it came to type II errors.

- **Two Year Forward Model**:
  The results attained for this forecast period varied with the type classifier employed. Overall, kNN outperformed KRR for this forecast period. The overall error rate of 12.3% for the kNN models was good if one considers that Altman (1968) had a 28% error rate for the same forecast period.

In terms of feature subsets, it appears as if Optimal Feature Subset A for the 2-Yr Fwd model had a better type I error performance, while Optimal Feature Subset B did better in reducing type II errors.

- **Three Year Forward Model**:
  KNN outperformed KRR in terms of type II errors, while both classifiers performed equally on type I errors. While KRR(C) achieved an overall error rate comparable to all the kNN models, KRR(A) and KRR(B) did not perform as well.

  The overall error rate of the 3-Yr Fwd model improved on that of the 2-Yr Fwd model. This was not expected as a model's ability to predict failure is expected to decrease as the time to failure increases. Such an outcome may have been as a result of the particular characteristics of the companies included in the sample. Alternatively, the poorer performance of the 2-Yr Fwd model may have been as a result of the PBIL runs not having identified the optimal feature subset for this forecast period, while it had done so for the 3-Yr Fwd model.

  Nevertheless, the results achieved with the 3-Yr Fwd models were good in relation to prior research (for example, Altman (1968) had a 52% error rate for the same forecast period).

  The only clearly superior feature subset identified within the set of 3-Yr Fwd models was Optimal Feature Subset C. This feature subset had the lowest type I error rate for both classifiers. The type II error rates achieved by the different feature subsets varied.

A superficial analysis of the above results indicated that KRR outperformed kNN as the quantity of data to be mapped increased. In other words, it performed better with a shorter forward forecast period - when there were a greater number of years of available data with which to perform the forecast.

As expected, the feature subset input into the classifier had a critical impact on the performance of that classifier.

21.4. RELATIVE MISCLASSIFICATION COSTS

When evaluating these models, the relative costs of a type I versus a type II errors needed to be considered. As Boritz & Kennedy (1995, 509) noted, it is unlikely that these errors are equally as costly. Many studies have been done on evaluating these relative costs. These studies, summarised and discussed in Chapter Ten, concluded that there is no generally accepted basis for trading off these different types of errors. Different stakeholders will have different perceptions of these relative costs.

This can be illustrated by considering the example of the differences in the concerns between the investors of a company and its management:

- **Investor**
  It is imperative to an investor that an investee company is not an imminent failure if it is considered to be a going concern investment (type I error). If, however, a company is deemed to be a failure and ends up succeeding (type II error), there is no nominal loss to the investor.

  As a result, the investor may be willing to tolerate a higher rate of type II error in order to have a lower type I error rate.

- **Management**
  In contrast, Boritz & Kennedy noted that management may prefer a lower rate of type II errors in order to avoid the "self-fulfilling prophecy" of a false failure signal, which may result in the market reacting accordingly.

21.4.1. Measuring Misclassification Costs

Due to the uncertainty of relative costs, misclassification costs were calculated under three assumptions. Cost ratios of 20:1, 1:1 and 1:20 (for (type I):(type II) costs) were used in assessing the relative performance of the models constructed. These ratios were selected as they have been used often in the published corporate failure research reviewed in this study (Boritz & Kennedy, 1995; Etheridge & Sriram, 1997).

Misclassification costs were calculated using the following formula:

$$\frac{(TypeIError \times \%FailedFirms \times CostRatio) + (TypeIIError \times \%NonFailedFirms)}{(\%FailedFirms \times CostRatio) + \%NonFailedFirms} \quad (21.1.)$$

21.4.2. ANALYSIS OF MODEL MISCLASSIFICATION COSTS

	20:1	1:1	1:20
1-Yr Fwd Model	KRR(A) kNN(A)	KRR(A) kNN(A)	KRR(A) kNN(A)
2-Yr Fwd Model	KRR(A) kNN(A) kNN(B) KRR(B)	KRR(A) kNN(A&B) KRR(B)	kNN(B) KRR(B) KRR(A) kNN(A)
3-Yr Fwd Model	kNN(C), KRR(C) kNN(A&B), KRR (A&B)	kNN(A,B&C), KRR(C) KRR(A&B)	kNN(A,B) KRR(A&B) kNN(C), KRR(C)

Table 21.3. Ranking of corporate failure prediction models by misclassification cost (Optimal feature subset in brackets)

Table 21.3. ranks the various prediction models based on their misclassification costs as presented in Table 21.1. Using the information in these two tables, the following was noted:

- The rankings of the models based on the 20:1 and 1:20 cost ratios were almost exactly the same as the rankings in Table 21.2., where rankings were based on type I and type II errors, respectively. This was to be expected, as the 20:1 and 1:20 cost ratios make their respective error types more costly and, hence, accentuate the impact of having made such a type of error.

- The only changes in relative performance in this regard related to the type II error rates of the 1:20 cost ratio rankings. Where two models had an equal performance in terms of type II error rates (per Table 21.2.), the incorporation of the type I error rate through the cost ratio resulted in one model outperforming another based on a greater type I error rate (per Table 21.3.). This ability to measure an error type, while still incorporating a degree of cost relating to the other error type, is a key advantage to using misclassification costs in the measurement of model performance.

  These ranking changes are as follows:

  - **1-Yr Fwd Model**: Where KRR(A) and kNN(A) had equal type II error rates, the higher type I error rate of the kNN(A) model resulted in it underperforming KRR(A) based on a 1:20 cost ratio.

- **2-Yr Fwd Model**: For the same reason as above, KRR(A) outperformed kNN(A) based on a 1:20 cost ratio.
- **3-Yr Fwd Model**: Once again, for the same reason as above, KRR(A) and KRR(B) outperformed KRR(C) and kNN(C).

- None of the type I error rate rankings changed with the incorporation of a nominal type II error cost when using the 20:1 cost ratio assumption. This may have been because many of the type I errors were significantly different from one another. Therefore, the type II error cost did not have an impact on the rankings of these models. It may also have been as a result of the negligible difference in type II costs between competing models.

- In terms of the impact of using different feature subsets, the following was noted:
    - **2-Yr Fwd Model:** Optimal Feature Subset A had a clear advantage over Optimal Feature Subset B when a higher cost was attributed to type I errors. The converse also held true.
    - **3-Yr Fwd Model**: Optimal Feature Subset C had a clear advantage over both Optimal Feature Subsets A and B when a higher cost was attributed to type I errors. Optimal Feature Subsets A and B performed better than C when type II errors were more costly. However, there was no noticeable difference in the performance between Optimal Feature Subsets A and B.

Once again, with the incorporation of misclassification costs into the evaluation of the models, KRR appeared to outperform kNN as the quantity of data to be mapped increased. This was the same for all the cost ratio assumptions that were tested.

Once again, as was expected, the feature subset input into the classifier had a critical impact on the performance of that classifier.

21.5. EVALUATING THE MISCLASSIFIED COMPANIES

Appendix I presents each model's predicted classes for each company. Note that these companies are a subset of the companies listed in Appendix D, as all companies with only two years of available data were excluded (this is explained in detail in Section B).

The following discussion evaluates the companies that were misclassified (highlighted in Appendix I). A detailed statistical assessment of the misclassified companies fell outside of the scope of this study and is dealt with as an area for further research in Chapter 23.

21.5.1. Analysis of Misclassification of "Borderline" Cases

A "reduced sample" database, excluding all cases defined as "borderline", was constructed as per Chapter Sixteen. These borderline cases were those failed companies that had met only a single failure criterion (defined in Chapter Thirteen) and those non-failed companies that had met four of the five failure criteria. Fifteen such failed cases and only a single non-failed case were identified (they can be identified using the information provided in Appendix D).

The following table summarises what proportion of these borderline cases were predicted correctly by **all** the models for each forecast period.

	Proportion of "borderline" failures classified correctly by all models	Proportion of "borderline" non-failures classified correctly by all models
1-Yr Fwd Model	93.33%	100%
2-Yr Fwd Model	73.33%	100%
3-Yr Fwd Model	80.00%	100%

Table 21.4. The proportion of "borderline" companies that have been accurately classified by all the constructed models for each failure forecast period

This analysis indicated that the models performed well in predicting the failure of those "borderline" companies. In fact, the 1-Yr Fwd models achieved a greater forecast accuracy on these cases than on the remaining sample.

The results presented here were also encouraging because the borderline non-failed company, even though it had met 80% of the failure criteria, was still classified correctly. This indicated that, when the "poor-performance" criteria were not combined with a significant change in the form of the existence of the company (as described in Chapter Thirteen), the criteria alone did not indicate that the said company had failed. This is what was conjectured when corporate failure was defined for the purposes of the selection of the sample for this study.

21.5.2. Potential for a Committee System Approach

	% (no.) co's not misclassified by any model	% (no.) co's not misclassified by at least one model
1-Yr Fwd Model		
- Failed Co.	90.16% (55)	91.80% (56)
- Non-Failed Co.	91.80% (56)	91.80% (56)
- Total	90.98% (111)	91.80% (112)
2-Yr Fwd Model		
- Failed Co.	83.61% (51)	98.36% (60)
- Non-Failed Co.	77.05% (47)	91.80% (56)
- Total	80.33% (98)	95.08% (116)
3-Yr Fwd Model		
- Failed Co.	86.89% (53)	90.16% (55)
- Non-Failed Co.	86.89% (53)	93.44% (57)
- Total	86.89% (106)	91.80% (112)

Table 21.5. Proportion of companies not misclassified by all and at least one of the prediction models constructed for each forecast period

There were two, four and six different prediction models constructed for the one, two and three year forward forecast periods, respectively. The above table summarises the proportion of companies not misclassified by any and at least one model for each forecast period.

The 2-Yr and 3-Yr Fwd models experienced a significant increase in the classification rate from column one to column two in the above table. This can be interpreted as follows: if one could select, *ex post*, which model to use for predicting each company for each forecast period, one could then achieve the results in the final column of the above table.

While such a scenario is not realistic, it does indicate that a committee system of classifiers may be an approach for improving classification accuracy. Such a method would involve constructing multiple models, using different feature subsets and/or different classification techniques, and then predicting the company classifications using each model. The final classification assigned to a particular case would then be a simple majority vote or weighted average of the predictions of each individual model for that particular case.

Preliminary "majority vote" models constructed using the data in this study indicated the potential for such an improvement in performance. This has been included as an area for further research in Chapter 23.

21.5.3. Feature Subset Classification Trends in the Misclassifications

It is clear from a brief inspection of Appendix I that a company misclassified by one model was, generally, also misclassified by at least one other model in that same forecast period. This may have been due to one of, or a combination of, the following factors:

- If the **input vector** of a company does not distinguish it correctly, then no classifier will be able to correctly predict the classification of such a company. As a result, the probability of the misclassification of a company, already misclassified by one model, may be greater if both models use the same input vector.

  Possible examples of this scenario are illustrated by the misclassifications of N27, N38, N42, N43, N63, F1, F6, F9, F38, F64 and F68 for the 2-Yr Fwd models, and by the misclassifications of N35, N64, F8 and F33 for the 3-Yr Fwd models. In all these instances, the companies was misclassified when both classifiers used the same feature subset for forecasting.

- There may be situations in which a particular **classification technique** maps the input data in a manner that does not allow for the correct classification of a company.

  There is no clear example in Appendix I of a situation in which a classifier misclassified a company using all feature subsets available to it, while the other classification method predicted that company accurately with all the same feature subsets. However, there are situations where, using the same feature subset, one classification technique was able to predict a company class accurately where the other technique did not. For example, KRR(A) correctly classified F11, while kNN(A) did not.

- Finally, the situation may arise where a company simply cannot be classified because it is **incorrectly assigned** to either the failed or non-failed class.

SECTION E: IN CONCLUSION

CHAPTER 22
CONCLUSION

The research objective of this study was to construct an empirical model for the prediction of corporate failure in South Africa through the application of machine learning techniques using information generally available to investors.

22.1.1. Summary of Model Construction

This objective was achieved by:

- Reviewing empirical corporate failure research and applying the relevant concepts tested and motivated in these studies.
- Defining corporate failure for the purposes of this study, taking into account the objectives of potential users, as well as the geographic location, data availability and classification methods employed.
- Selecting a sample of failed and non-failed companies from the JSE Securities Exchange over the period January 1996 to June 2003 using selection criteria based on the said definition.
- Collecting at least two years and a maximum of three years of commonly used information for each company from a source that is accessible to the majority of South African investors.
- Processing this financial data into a format that suited the methodologies employed and best helped to cluster the data.
- Applying Population-Based Incremental Learning (PBIL) to the feature subset selection problem.
- Using the optimal feature subsets to construct corporate failure prediction models for a one, two and three year forward forecast periods using k-Nearest Neighbours (kNN) and Kernel Ridge Regression (KRR).
- Evaluating the models constructed.

22.1.2. Implementation of PBIL

PBIL was successfully applied to the feature subset selection problem. The iterative process appears to have functioned as intended.

In addition, the algorithm was able to identify a number of different feature subsets that had equally good data separability. This was consistent with the expectation that different combinations of financial ratios can bear similar information content with regards to assessing corporate financial health. In fact, the way the PBIL algorithm randomly search a solution space and then focuses on a particular optimum within that space, was well adapted to this problem.

Thornton's Separability Index performed well in identifying those feature subsets that performed well when input into the proximity-based classifiers employed in this study.

22.1.3. IMPLEMENTATION OF kNN AND KRR

Two proximity-based, each using a one, two and three year forward forecast period, were applied to each of the optimal feature subsets identified by the PBIL algorithm.

The k-Nearest Neighbour classifier used was a simple 1NN. "One" was clearly identified as the optimum value for k.

The parameters of the Kernel Ridge Regression algorithm were identified on a trial-and-error basis. Gamma, the regularisation parameter, was found to have a negligible impact on the classification accuracy of the model. This was interpreted as meaning that the data sets were both numerically well-conditioned and that the data clusters had a significant overlap. The sigma (or kernel width) was found to have a significant impact on the accuracy of the KRR models. It was tested across an "optimal range" of values that was determined on a trial-and-error basis.

There was no classifier that clearly outperformed another. However, it did appear as if the KRR was able to better map the distinction between classifications on the larger data sets. A detailed comparison of the two classifiers is presented in the preceding chapter.

22.1.4. FINAL MODELS RESULTING FROM THE CONSTRUCTION PROCESS

The end results of this construction process were one, two and three year forward corporate failure prediction models, each using one, two and three optimal feature subsets, respectively, with both a kNN and KRR classifier. The results of these twelve models are summarised and evaluated in the preceding chapter.

With over 90% prediction accuracy for the one and three year forward forecast periods, as well as a prediction accuracy of no less than 86% for the two year forecast period, this construction process was considered to have been successful.

CHAPTER 23

RELATED TOPICS FOR FURTHER RESEARCH

Throughout this report areas for further research have been identified. As these areas did not fall within the scope of the research objective of this study, they are presented in this chapter as topics for future researcher.

23.1. COMPUTATIONAL RESOURCES

A key restriction in the use of iterative machine learning techniques is the amount of time it takes to run a particular algorithm on a particular set of data. The trade-off between learning speed and model accuracy is an important one in real-world applications. Slightly reduced accuracy may be advantageous if it allows for a rapid increase in classification time, especially where large amounts of data need to be processed. Therefore, creating ways in which to reduce the learning speed, in terms of both feature subset selection and inductive learning, is a potentially valuable research problem that could be explored.

23.2. INFORMATION CONTENT OF FINANCIAL DATA

In Chapter Eighteen it is noted that when the PBIL algorithm was run on the full set of data multiple times, a number of different feature subsets, with equal separability, were identified. It is also noted in Chapter Five, and investigated superficially in Chapter Seventeen, how the interpretation of a ratio can be dependent on with which other ratios it is analysed.

The feature subsets selected as optimal for the classifiers in this study were not analysed in any detail.

Much research has been performed in the literature on what factors contribute to failure and how these factors can best be assessed using financial statement information. In Chapter Eighteen, ten feature subsets, all with high Separability Index values, were selected for each forecast period and have been disclosed in Appendix H. Suggested analysis on such subsets include:

- Identifying which features were not selected or which features were selected only a few times for a feature subset that had good data separability.

- Identifying which features were selected on all or most PBIL runs.
- Identifying how the types of features that lead to good class separability changed as the time to failure decreased (i.e. a comparison between the types of features selected in the one, two and three year forward forecast feature subsets)
- Assessing how, within a single forecast period, the features selected for each subset differed across the years prior to date of failure.
- Assessing the information content of a feature in how it was combined with different features in its subset to explain the difference between failure and non-failure.

23.3. ALTERNATIVE APPROACHES FOR IMPROVING FEATURE SUBSET SELECTION

23.3.1. PRE-PROCESSING

In this study, the features were normalised before being input into the PBIL algorithm.

Cantu-Paz (2002, 2) suggested an additional step that can be performed prior to running the feature subset selection algorithm. Using Genetic Algorithms, one can attempt to extract *new* features by searching for a vector of numeric coefficients that can be used to linearly transform the original variables before subset selection. PBIL could also be used for this purpose.

23.3.2. REFINEMENTS TO THE PBIL ALGORITHM

Additional refinements that can be made to PBIL and tested for improved feature subset selection include:

- Baluja (1996) tested an alternative version of PBIL learning that not only pushes the probability vector towards the best solution, but also moves this vector away from the worst.
- Inza et al (1999) pushed the probability vector towards the best *M* solutions rather than the single best bitstring evaluated.
- Baluja (1996) proposed introducing parallel running PBIL algorithms that evaluate multiple probability vectors. A form of crossover can then also be introduced.
- Baluja and Caruana (1995, 8) proposed a variation on PBIL where the probability vector is incrementally updated as each new trial is generated - rather than updating it from only a few of the best trial solutions generated on each iteration.

23.3.3. Selection of PBIL Parameters

The parameters that need to be selected in order to run the PBIL algorithm are non-critical (Greene, 1996). As a result, the methods used in this study for determining the parameter values were crude and subjective. It was outside of the scope of this study to perform detailed research in order to find the optimal parameter values.

Detailed testing for determining these parameters, in order to improve the process, may be valuable and is an area for further research.

23.4. POTENTIAL FOR THE APPLICATION OF A COMMITTEE SYSTEM APPROACH TO MODEL CONSTRUCTION

A committee system involves constructing a number of different models (using different feature subsets and/or different classification techniques) and then taking some sort of average of each model in deciding the classification of an unseen case. Averaging can be performed on a simple "majority vote" basis or by another weighted method.

The potential for such an approach lies in the facts that:

- PBIL is able to identify numerous different feature subsets that have equally good separability but still explain the data in different ways; and
- KRR and kNN do not map the data in the same way - resulting in some companies being misclassified by one technique but not the other, even though they may use the same feature subset.

Therefore, classification accuracy may be improved by utilising the information from all feature subsets and classification techniques. A preliminary assessment of a committee system using all the models and feature subsets constructed in this study indicated that such an improvement may be attainable.

23.5. ANALYSIS OF THE MISCLASSIFIED COMPANIES

The companies misclassified by the different models are highlighted in Appendix I.

Further analysis of misclassifications in terms of such factors as industry, year of failure, feature subset, etc. may reveal information regarding the applicability of the models to different real-world problems. In addition, one can evaluate the impact of these factors on corporate failure and, hence, address another key area of corporate failure research – the study of the causes and symptoms of failure.

REFERENCES

Journal Articles

Ahn, B.S., Cho, S.S. & Kim, C.Y. (2000) "The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction". *Expert Systems with Applications*, 18, 65-74.

Alexander, S. (1949) "The Effect of Size of Manufacturing Corporations on the Distribution of the Rate of Return". *Review of Economics and Statistics*, August, 229-235.

Altman, E.I. (1968) "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *Journal of Finance*, 23, 4, 589-609.

Altman, E.I., Haldeman, R.G. & Narayanan, R. (1977) "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations". *Journal of Banking and Finance*, 1, 29-54.

Altman, E.I., Marco, G. & Varetto, F. (1994) "Corporate Distress Diagnosis: Comparisons using Linear Discriminant Analysis and Neural Networks (the Italian Experience)". *Journal of Banking and Finance*, 18, 3, 509-529.

Altman, E.J. (1984) "The success of business failure prediction models, an international survey". *J. of Banking and Finance*, 8, 171-198.

Atiya, A.F. (2001) "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results". *IEEE Transactions on Neural Networks*, 12, 4, 929-935.

Back, B., Laitinen, T. & Sere, K. (1996) "Neural Networks and Genetic Algorithms for Bankruptcy Predictions." *Expert System Applications*, 407-413.

Back, B., Oosterom, G., Sere, K. & Van Wezel, M. (1995) "Choosing the Best Set of Bankruptcy Predictors". *Alander*, 1000, 285-301.

Baluja, S. (1996) "Genetic Algorithms and Explicit Search Statistics". *Advances in Neural Information Processing Systems (NIPS)*, December, 1-7.

Bamiv, R., Agarwal, A. & Leach, R. (1997) "Predicting the Outcome Following Bankruptcy Filing: A three-state classification using neural networks". *Intelligent Systems in Accounting, Finance and Management*, 6, 177-194.

Barron, A.R. (1994) "Approximation and Estimation Bounds for Artificial Neural Networks". *Machine Learning*, 14, 115-133.

Battacharya, S. (1979) "Imperfect Information, Dividend Policy and the Bird in Hand Fallacy". *Bell Journal of Economics*, 10, 259-270.

Beaver, W. Kennelly, J.W. & Voss, W.M. (1968) "Predictive Ability as a Criterion for the Evaluation of Accounting Data". *The Accounting Review,* October, 675-683.

Beaver, W.H. (1966) "Financial Ratios as Predictors of Failure. Empirical Research in Accounting: Selected Studies". *Supplement to Journal of Accounting Research*, 5, 71-111.

Beaver, W.H. (1968) "Market Prices, Financial Ratios and Prediction of Failure". *Journal of Accounting Research*, Autumn.

Bell, T.B. (1997) "Neural Nets or the Logit Model? A Comparison of each Model's Ability to Predict Commercial Bank Failure". *Intelligent Systems in Accounting, Finance and Management*, 6, 249-264.

Blum, M. (1974) "Failing Company Discriminant Analysis". *Journal of Accounting Research*, 12, 1, 1-25.

Boritz, J.E. & Kennedy, D.B. (1995) "Effectiveness of Neural Network Types for Prediction of Business Failure". *Expert Systems with Applications*, 9, 4, 503-512.

Boritz, J.E., Kennedy, D.B. & De Miranda e Alburquerque, A. (1995) "Predicting Corporate Failure Using a Neural Network Approach". *Intelligent Systems in Accounting Finance and Management*, 4, 95-111.

Box, G.E.P. (1949) "A General Distribution Theory for a Class of Likelihood Criteria". *Biometrika*, 36.

Brockett, P.L., Cooper, W.W., Golden, L.L. & Pikatong, U. (1994) "A Neural Network Method for Obtaining an Early Waming of Insurer Insolvency". *The Journal of Risk and Insurance*, 61, 3, 402-424.

Bryant, S.M. (1997) "A Case-Based Reasoning Approach to Bankruptcy Prediction Modeling". *Intelligent Systems in Accounting, Finance and Management*, 6, 195-214.

Bulow, J. & Shoven, J. (1978) "The Bankruptcy Decision". *The Bell Journal of Economics*, 9, 2, 437-456.

Casey, M., McGee, V. & Stinkey, C. (1986) "Discriminating between Reorganised and Liquidated firms in Bankruptcy". *The Accounting Review,* April, 249-262.

Chatfield, C. (1993) "Neural Networks: Forecasting Breakthrough or Passing Fad?". *International J. of Forecasting*, 9, 1-3.

Chesser, D.L. (1974) "Predicting Loan Noncompliance." *The Journal of Commercial Banking Lending,* August, 28-38.

Coleman, K.G., Graettinger, T.J. & Lawrence, W.F. (1991) "Neural Networks for Bankruptcy Prediction: The Power to Solve Financial Problems". *AI Review*, 48-50.

Cooper, R.G. (1979) "The Dimensions of Industrial New Product Success and Failure". *Journal of Marketing,* 43, Summer, 93-103.

Court, P.W. & Radloff, S.E. (1990) "A Comparison of Multivariate Discriminant and Logistic Regression Analysis in the Prediction of Corporate Failure in South Africa". *De Ratione*, 4, 2, 11-15.

Court, P.W. & Radloff, S.E. (1994) "A Two-Stage Model for the Prediction of Corporate Failure in South Africa". *Investment Analysts Journal*, 38, Summer, 9-19.

Court, P.W. (1991) "An investigation into the significance of certain firm-specific non-financial variables in a failure prediction model". *De Ratione*, 5, 2, 3-15.

Cox, D.R. (1972) "Regression Models and Life Tables". *Journal of the Royal Statistical Society,* Series B, 187-220.

Crouhy, M., Galai, D. & Mark, R. (2000) "A Comparative Analysis of Current Credit Risk Models". *Journal of Banking and Finance*, 24, 59-117.

Crowley, J. & Hu. M. (1977) "Covariance Analysis of Heart Transplant Survival Data". *Journal of the American Statistical Association,* March, 27-36.

Cucker, F. & Smale, S. (2001) "On the Mathematical Foundations of Learning." *Bull. Amer. Math. Soc. (N.S.)*, 39, 1-49.

Davidson, R. & MacKinnon, J.G. (1984) "Convenient Specifications Test for Logit and Probit Models". *Journal of Econometrics*, 25, 3, 241-262.

De la Rey, J.H. (1981) "Finansiele verhoudingsgetalle en die voorspelling van finasiele (1981) mislukking be nywerheidsondernemings in die Republiek van Suid-Afrika". *Report E1*, Buro van Finansiele Analise, University of Pretoria.

Deakin, E.B. (1972) "A Discriminant Analysis of Predictors of Business Failure". *J. of Accounting Research*, Spring, 167-179.

Dimitras, A., Zanakis, S. & Zopounidis, C. (1996) "A Survey of Business Failure with an Emphasis on Prediction Methods and Industrial Applications". *European Journal of Operational Research*, 90, May, 487-513.

Easterbrook, F. (1990) "Is Corporate Bankruptcy Efficient?", *Journal of Financial Economics*, 27, 411-417.

Eisenbeis, R.A. (1977) "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics". *J. of Finance*, 32, 3, 875-900.

Eisenbeis, R.A., Gilbert, G.G. & Avery, R.B. (1973) "Investigating the Relative Importance of Individual Variables and Variable Subsets in Discriminant Analysis". *Communications in Statistics*, 2, 3, 205-219.

El Hennawy, R.H.A. & Morris, R.C. (1983) "The Significance of Base Year in Developing Failure Prediction Models". *J of Business Finance & Accounting*, 10, 2, 209-223.

Elam, R.J. (1975) "The Effect of Lease Data on the Predictive Ability of Financial Ratios". *The Accounting Review*, 5, 1, 25-43.

Els, N. (1996(a)) "The A-Z of Corporate Failure. Part 1: Z-Scores". *Accountancy SA*, January.

Els, N. (1996(b)) "The A-Z of Corporate Failure. Part 1: Z-Scores". *Accountancy SA*, February.

Etheridge, H.L. & Sriram, R.S. (1997) "A Comparison of the Relative Costs of Financial Distress Models: Artificial Neural Networks, Logit and Multivariate Discriminant Analysis". *Intelligent Systems in Accounting, Finance and Management*, 6, 235-248.

Fletcher, D. & Goss, E. (1993) "Forecasting with Neural Networks: An Application using Bankruptcy Data". *Information and Management*, 24, 159-167.

Flynn, D.K. (1987) "Perceptions Regarding the Sources of Financial Information for South African Institutional Investors". *De Ratione*, 1, 2-7.

Frank, R.E., Massy, W.F. & Morrison, G.D. (1965) `Bias in Multiple Discriminant Analysis'. *Journal of Marketing Research*, 2, August, 220-258.

Frydman, H., Altman, E.I. & Kao, D. (1985) "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress". *Journal of Finance*, 269-291.

Girosi, F. (1998) "An Equivalence between Sparse Approximation and Support Vector Machines". *Neural Computation*, 10, 6, 1455-1480.

Hambrick, D. & D'Aveni, R. (1988) "Large Corporate Failures as Downward Spirals". *Administrative Science Quarterly*, 33, 1-23.

Hansen, J.V. (1998) "Comparative Performance of Backpropogation Network Designed by Genetic Algorithms and Heuristics". *International J. of Intelligent Systems in Accounting Finance and Management*, 7, 69-79.

Hofstedt, T.R. (1972) "Some Behavioural Parameters of Financial Analysis". *The Accounting Review*, 47, October, 679-692.

Huang, Xiang-Ning & Li, Bai-Bing (1991) "A New Discriminant Technique: Bayes-Fisher Discrimination". *Biometrics*, 47, 741-744.

Hubbard, G. (1994) "Investment under Uncertainty: Keeping One's Options Open". *Journal of Economic Literature*, 32, 4, 1816-1831.

Ijiri, Y. (1967) *The Foundations of Accounting Measurement*. Prentice-Hall, Englewood Cliffs.

Jarrow, R. & Turnbull, S. (1995) "The Pricing and Hedging of Options on Financial Securities Subject to Credit Risk". *Journal of Finance*, 50, 53-85.

John, K. & Williams, J. (1985) "Dividends, Dilution and Taxes: A Signaling Equilibrium". *Journal of Finance*, 40, 1053-1070.

Johnson, C.G. & Altman, E.I. (1970) "Ratio Analysis and the Prediction of Firm Values". *Journal of Finance*, December, 1166-1172.

Joy, O. & Tollefson, J. (1975) "On the Financial Applications of Discriminant Analysis". *Journal of Financial and Quantitative Analysis*, 10, 5, 723-741.

Kaastra, I. & Boyd, M. (1996) "Designing a Neural Network for Forecasting Financial and Economic Time Series". *Accounting and Business Research*, 29, 3, 211-216.

Kaski, S., Sinkkonen, J. & Peltonen, J. (2001) "Bankruptcy Analysis with Self-Organising maps in Learning Metrics". *IEEE Transactions on Neural Networks*, 12, 4, 936-947.

Koh, H.C. & Tan, S.S. (1999) "A Neural Network Approach to the Prediction of Going Concern Status". *Accounting and Business Research*, 29, 3, 211-216.

Korobow, L. & Stuhr, D. (1985) "Performance Measurement of Early Warning Models". *Journal of Banking and Finance*, 9, 267-273.

Lachenbruch, P. & Mickey, M.R. (1968) "Estimation of Error Rates on Discriminant Analysis". *Technometrics*, 10, 1.

Lachenbruch, P. (1967) "An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis". *Biometrics*, December.

Lanquillon, C. (1999) "Dynamic Aspects in Neural Classification". *International J. of Intelligent Systems in Accounting Finance and Management*, 8, 281-296.

Lau, A.H. (1987) "A Five-state Distress Prediction Model". *Journal of Accounting Research*, 25, 1, 127-138.

Lawrence, E.C. (1983) "Reporting Delays for Failed Firms". *Journal of Accounting Research*, 21, 2, 606-628.

Lee, K, Han, I. & Kwon, Y. (19996) "Hybrid Neural Network Models for Bankruptcy Predications". *Decision Support Systems*, 18, 63-72.

Lee, K.C., Han, I. & Kwon, Y. (1996) "Hybrid Neural Network Models for Bankruptcy Predictions". *Decision Support Systems*, 18, 63-72.

Lennox, C. (1999) "Identifying Failing Companies: A Re-evaluation of the Logit, Probit and

DA Approaches". *Journal of Economics and Business*, 51, 347-364.

Lennox, C.S. (1999) "The Accuracy and Incremental Information Content of Audit Reports in Predicting Bankruptcy". *J. of Business Finance & Accounting*, 26, 5, 757-778.

Leshno, M. & Spector, Y. (1996) "Neural Network Prediction Analysis: The Bankruptcy Case". *Neurocomputing*, 10, 125-147.

Libby, R. (1975) "Accounting Ratios and the Prediction of Failure: Some Behavioural Evidence". *Journal of Accounting Research*, March.

Lo, A.W. (1986) "Logit versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcies". *Journal of Econometrics*, 31, 2, 151-178.

Longstaff, F. & Schwartz, E. (1995) "A Simple Approach to Valuing Risky Fixed and Floating Rate Debt". *Journal of Finance*, 50, 789-819.

Looney, S.W., Wansley, J.W. & Lane, W.R. (1986) "An Application of the Cox Proportional Hazards Model to Bank Failure". *Journal of Banking and Finance*, 10, 511-531.

Looney, S.W., Wansley, J.W. & Lane, W.R. (1989) "An Examination of Misclassifications with Bank Failure Prediction Models". *Journal of Economics and Business*, 41, 327-336.

Martin, D. (1977) "Early Warnings of Bank Failure". *Journal of Banking and Finance,* 249-267.

Mensah, Y.M. (1984) "An Examination of the Stationarity of Multivariate Bankruptcy Prediction Models: A Methodological Study". *Journal of Accounting Research,* 22, 380-395.

Merton, R. (1974) "On the pricing of Corporate Debt: The Risk Structure of Interest Rates". *Journal of Finance*, 29, 449-470.

Meyer, D., Leich, F. & Homik, K. (2003) "The support vector machine under test." *Neurocomputing*, 55, 1-2, 169-186.

Miller, M. & Modigliani, F. (1961) "Dividend Policy, Growth and the Valuation of Shares". *Journal of Business*, 34, 411-433.

Mitchell, D.W. (1976) "Off the Balance Sheet - Way Off!". *Journal of Commercial Banking Lending,* August, 2-8.

Müller, K., Mika, S., Rätsch, G., Tsuda, K. & Schölkopf, B. (2001) "An Introduction to Kernel-Based Learning Algorithms". *IEEE Transactions on Neural Networks*, 12, 2, 181-202.

Myers, S. (1977) "Determinants of Corporate Borrowing". *Journal of Financial Economics*, 5, 2, 147-175.

Niyogi, P. & Girosi, F. (1996) "On the Relationship between Generalization Error, Hypothesis Complexity, and Sample Complexity for Radial Basis Functions". *Neural Computing*, 8, 819-842.

Norton, C.L. & Smith, R.E. (1979) "A Comparison of General Price Level and Historical Financial Statements in the Prediction of Bankruptcy". *The Accounting Review*, 54, 72-87.

Ohlson, J.A. (1980) "Financial Ratios and the Probabilistic Prediction of Bankruptcy". *Journal of Accounting Research*, Spring, 109-131.

O'Leary, D.E. (1998) "Using Neural Networks to Predict Corporate Failure". *International J. of Intelligent Systems in Accounting Finance and Management*, 7, 187-197.

Peel, M.J., Peel, D.A. & Pope, P.F. (1986) "Predicting Corporate Failure - Some results for the UK Corporate Sector". *Omega*, 14, 2, 5-12.

Piramuthu, S., Raghavan, H. & Shaw, M. (1998) "Using Feature Construction to Improve the Performance of Neural Networks". *Management Science*, 44, 416-430.

Piramuthu, S., Shaw, M. & Gentry, J.A. (1994) "A Classification Approach using Multi-Layered Neural Networks". *Decision Support Systems*, 11, 509-525.

Platt, H., Platt, M. & Pederson, J. (1994) "Bankruptcy Discrimination with Real Variables". *Journal of Business, Finance and Accounting*, 21, 4, 491-510.

Platt, H.D. & Platt, M.B. (1990) "Developing a Stable Class of Predictive Variables: The Case of Bankruptcy Prediction". *Journal of Business, Finance and Accounting*, 17, 1, 1183-1194.

Platt, H.D. (1989) "The Determinants of Industry Failure". *Journal of Economics and Business*, 41, 107-126.

Poggio, T. & Smale, S. (2003) "The Mathematics of Learning: Dealing with Data". *Notice of the AMS*, 50, 5, 537-544.

Ramamoorti, S., Bailey, A.D. & Traver, R.O. (1999) "Risk Assessment in Internal Auditing: A Neural Network Approach". *International J. of Intelligent Systems in Accounting Finance and Management*, 8, 159-180.

Richardson, F.M. & Davidson, L.F. (1983) "An Exploration into Bankruptcy Discriminant Model Sensitivity". *Journal of Business Finance and Accounting*, 195-207.

Rose, P.S., Andrews, W.T. & Giroux, G.A. (1982) "Predicting Business Failure: A Macroeconomic Perspective". *Journal of Accounting, Auditing and Finance*, 6, 20-31.

Ruiz, A. & Lopez-de-Tereul, P.E. (2001) "Nonlinear Kernel-Based Statistical Pattern Analysis". *IEEE Transactions on Neural Networks*, 12, 1, 16-32.

Ryan, P.A., Besley, S. & Lee, H.W. (2000) "An Empirical Analysis of Reactions to Dividend Policy Changes for NASDAQ Firms". *Journal of Financial and Strategic Decisions*, 13, 1, 35-44.

Salchenberger, L. Cinar, E. & Lash, N. (1992) "Neural Networks: A New Tool for Predicting Thrift Failure". *Decision Sciences*, 23, 899-916.

Scott, J. (1981) "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models". *Journal of Banking and Finance*, 5, 317-344.

Sharda, R. & Wilson, R.L. (1996) "Neural Network Experiments in Business-Failure Forecasting". *International Journal of Computational Intelligence and Organizations*, 1, 2, 107-117.

Sharma, S. & Mahajan, V. (1980) "Early Warning Indicators of Business Failure". *Journal of Marketing*, 44, Fall, 80-89.

Slowinski, R. & Zopounidis, C. (1995) "Application of the Rough Set Approach to Evaluation of Bankruptcy Risk". *Intelligent Systems in Accounting, Finance and Management*, 4, 27-41.

Stiglitz, J.E. (1972) "Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Takeovers". *Bell Journal of Economics and Management Science*, Autumn, 458-482.

Stone, M. (1974) "Cross-Validatory Choice and Assessment of Statistical Predictions". *Journal of Royal Statistical Society*, B 36 (1), 111 - 147.

Stone, M. (1978) "Cross-Validation: A Review". *Math. Operationsforsch. Statist. Ser. Statistics*, 9, 1, 127-139.

Sutton, S.G., Young, R. & McKenzie, P. (1994) "An Analysis of Potential Legal Liability Incurred Through Audit Expert Systems". *Intelligent Systems in Accounting, Finance and Management*, 4, 191-204.

Swinney, L. (1999) "Consideration of the Social Context of Auditors' Reliance on Expert System Output during Evaluation of Loan Loss Reserves". *International J. of Intelligent Systems in Accounting Finance and Management*, 8, 199-213.

Taffler, R.J. & Citron, D.B. (2001) "Ethical Behaviour in the U.K. Audit Profession: The Case of the Self-Fulfilling Prophecy Under Going Concern Uncertainties". *Journal of Business Ethics*, 29, 353-363.

Tam, K. & Kiang, M. (1992) "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions". *Management Science*, 38, 416-430.

Tam, K. (1991) "Neural Network Models and the Prediction of Bank Bankruptcy". *Omega*, 19, 429-445.

Treacy, W. & Carey, M. (2000) "Credit Risk Rating at Large US Banks". *Journal of Banking and Finance*, 24, 167-201.

Trigueiros, D. & Taffler, R. (1996) "Neural Networks and Empirical Research in Accounting". *Accounting and Business Research*, 26, 4, 347-355.

Van Gestel, T., Suykens, J.A., Baestens, D., Lambrechts, A., Lanckriet, G., Vandaele, B., De Moor, B & Vandewalle, J. (2001) "Financial Time Series Prediction Using Least Squares Support Vector Machines within the Evidence Framework". *IEEE Transactions on Neural Networks*, 12, 4, 809-820.

Wahba, G. & Wold, S. (1975) " A Completely Automatic French Curve: Fitting spline functions by cross-validation". *Communications in Statistics*, Series A, 4, 1, 1-17.

Wahba, G. (1990) "Splines Models for Observational Data" *Series in Applied Mathematics*, vol. 59, SIAM, Philadelphia, P.A.

Welch, O.J., Reeves, T.E. & Welch, S.T. (1998) "Using Genetic Algorithm-Based Classifier System for Modeling Auditor Decision Behaviour in a Fraud Setting". *International J. of Intelligent Systems in Accounting Finance and Management*, 7, 173-186.

Whittred, G.P. & Zimmer, I. (1984) "Timeliness of Financial Reporting and Financial Distress". *Accounting Review*, April, 287-296.

Wiggins, N. (1973) "Individual Differences in Human Judgement: A Multivariate Approach". *Human Judgement and Social Interaction*, Eds. Rappaport, L. & Summers, D.A., Holt, Rinehart and Winstin, New York.

Wilcox, J. (1973) "A Prediction of Business Failure using Accounting Data". *Journal of Accounting Research*, Supplement, 11, 163-179.

Wilcox, J. (1976) "The Gamblers Ruin Approach to Business Risk". *Sloan Management Review*, Autumn.

Wilson, R.F. & Sharda, R. (1994) "Bankruptcy Prediction using Neural Networks". *Decision Support Systems*, 11, 545-577.

Wong, B., Bodnovich, T. & Selvi, Y. (1997) "Neural Network Applications in Business: A Review and Analysis of the Literature". *Decision Support Systems*, 19, 301-320.

Wruck, K. (1990) "Financial Distress, Reorganisation and Organisational Efficiency". *Journal of Financial Economics*, 27, 2, 419-444.

Wuthrich, B. (1997) "Discovering Probabilistic Decision Rules". *Intelligent Systems in Accounting, Finance and Management*, 6, 269-277.

Yatchew, A. & Griliches, Z. (1985) "Specification Error in Probit Models". *The Review of Economics and Statistics*, 66, 1, 134-139.

Zavgren, C.V. (1985) "Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis". *Journal of Business Finance and Accounting*, 12, 1, 19-45.

Zhang, G., Hu, M. & Patuwo, B. (1999) "Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis". *European Journal of Operations Research*, 116, 16-32.

Zmijewski, M.E. (1984) "Methodological Issues Related to the Estimation of Financial Distress Prediction Models". *Journal of Accounting Research*, Supplement, 22, 59-86.

Books

Altman, E.I., Avery, R.B., Eisenbeis, R.A. & Sinkey, J.F. (1981) *Application of Classification Techniques in Business, Banking and Finance, Contemporary Studies in Economic and Financial Analysis,* Vol3. JAI Press, Greenwich, CT.

Altman, E.J. (1983) *Corporate Financial Distress. A Complete Guide to Prediction, Avoiding and Dealing with Bankruptcy.* Wiley, New York, First Edition.

Arbib, M.A. (1995) *The Handbook of Brain Theory and Neural Networks.* The MIT Press: USA.

Argenti, J. (1976) *Corporate Collapse: the causes and symptoms.* McGraw-Hill, London, First Edition

Bellman, R. (1961) *Adaptive Control Processes: A guided tour.* New Jersey: Princeton University Press.

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition.* Oxford University Press, Oxford.

Blackman, M.S. (1996) *Laws of South Africa,* Ed. Joubert, W.A. Vol.4, No. 3, Butterworths, Durban.

Bodie, Z., Kane, A. & Marcus, A.J. (1996) *Investments,* 3rd Ed. McGraw-Hill Companies, Inc., Boston.

Bow, S. (1984) *Pattern Recognition: Application to Large Data-Set Problems.* Marcel Dekker, Inc., New York, USA.

Correia, C., Flynn, D., Uliana, E. & Wormald, M. (2000) *Financial Management (4th Edition).* Juta & Co. Ltd, Cape Town, South Africa.

Cristianini, N. & Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods.* Cambridge University Press, Cambridge.

Dixit, A. & Pindyck, R. (1994) *Investment under Uncertainty.* Princeton University Press, Princeton.

Donaldson, G. (1969) *Strategy for Financial Mobility.* Boston: Division of Research, Graduate School of Business Administration, Harvard University.

Duda, R.O. & Hart, P.E. (1973) *Pattern Classification and Scene Analysis.* Wiley, New York 1973

Fisher, R. (1952) *Contributions to Mathematical Statistics.* Wiley, New York.

Haykin, S. (1994) *Neural Networks: A Comprehensive Foundation.* MacMillan: New York.

Kerling, M. (1996) "Corporate Distress Diagnosis - An International Comparison". In: *Neural Networks in Financial Engineering,* (Eds.) Refenes, A.P.N., Abu-Mostafa, Y., Moody, J. & Weigend, A., World Scientific, Singapore, 407-422.

Platt, H.D. (1985) *Why Companies Fail: Strategies for detecting, avoiding and profiting from bankruptcy,* First Ed. Heath and Company, Lexington.

Poddig, T. (1995) "Bankruptcy Prediction: A Comparison with Discriminant Analysis". In: *Neural Networks in Capital Markets,* (Ed.) Refenes, A.P.N., Wiley, Chichester, 311-324.

Rahimian, E., Singh, S., Thammachote, T. & Virmani, R. (1993) "Bankruptcy Prediction by Neural Network". in: *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance,* (Eds.) Trippi, R. & Turban, E., Probus, Chicago. 159-176.

Rand, A. (1961) *For the New Intellectual: The Philosophy of Ayn Rand.* The New York American Library, New York.

Rifkin, R., Yeo, G. & Poggio, T. (2003) "Regularised Least-Squares Classification". In: *Advances in Learning Theory: Methods, Model and Applications,* NATO Science Series III: Computer and Systems Sciences, VIOS Press, Amsterdam, (Eds.) Suykens, Horvath, Basu, Micchelli and Vandewalle, 190, Chapter 7, 131-154

Scott, D.W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualisation.* New York: John Wiley.

Schalkoff, R. (1992) *Pattern Recognition: Statistical, Structural and Neural Approaches.* John Wiley & Sons, Inc., Canada.

Schumpeter, J. (1934) *The Theory of Economic Development,* Cambridge, MA: Harvard Business Press.

Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B. & Vandewalle, J. (2002) *Least Squares Support Vector Machines.* World Scientific Publishing Company.

Tikhonov, A.N. & Arsenin, V.Y. (1977) *Solutions of Ill-Posed Problems.* W.H. Winston

Van Den Honert, R. (1997) *Intermediate Statistical Methods for Business and Economics*. UCT Press (Pty) Ltd, Rondebosch, South Africa.

Van Horne, J.C. (1986) *Financial Management and Policy,* Seventh Ed. Prentice/Hall, Englewood Cliffs.

Vapnik, V. (1998) *Statistical Learning Theory*. Wiley, New York.

Zirilli, J.S. (1997) *Financial Prediction using Neural Networks.* London: International Thomson Computer Press.

Unpublished Works

Arron, J.C. (1994) *The Use of Neural Networks in Predicting Corporate Failure.* MBA: University of the Witwatersrand.

Back, B., Laitinen, T., Sere, K. & Van Wezel, M. (1996) *Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms.* Technical Report No 40, Turku Centre for Computer Science, September.

Court, P.W. (1993) *A Combination of a Stationary and Non-Stationary Model to Predict Corporate Failure in South Africa.* PhD: Rhodes University.

Daya, K. (1977) *Financial Ratios as Predictors of Corporate Failure in South Africa.* MBA: University of the Witwatersrand.

Dorsey, R.E., Edmister, R.O. & Johnson, J.D. (1994) *Bankruptcy Prediction Using Artificial Neural Networks.* University of Mississippi School of Business (unpublished).

Hahn, B.D. (1997) *Essential Matlab for Scientists and Engineers.* Department of Mathematics and Applied Mathematics, University of Cape Town, South Africa.

Henery, R.J. (1994) "Classification". In: *Machine Learning, Neural and Statistical Classification*, Statlog Project, (Eds.) Michie, D., Spiegelhalter, D.J. & Taylor, C.C., Chapter 2, 6-16.

Greene, J.R. (2001b) Soft Computing, EEE496S Lecture Notes.

Greene, J.R. (2002) Soft Computing, EEE496S Lecture Notes.

Greene, J.R. (2003) Soft Computing, EEE496S Lecture Notes.

Greene, J.R. (2004) Private discussions.

Immelman, D.A. (1980) *A Multi-Variate Approach Employing Stock Market Returns to the Prediction of Corporate Failure.* MBA: University of the Witwatersrand.

Inza, I., Merino, M., Larrañaga, P., Quiroga, J., Sierra, B. & Girala, M. (1999) "Feature Subset Selection by Population-Based Incremental Learning. A case study in the survival of cirrhotic patients treated with TIPS". *Technical Report No EHU-KZAA-IK-1/99*, University of the Basque Country, Spain.

Le Roux, J. (1980) *A Study in the Use of Discriminant and Factor Analysis as Aids in the Prediction of Corporate Failure.* MBA: University of Stellenbosch.

Merks, D.G. (1986) *Non-Financial Ratios as Predictors of Corporate Failure.* MBA: University of the Witwatersrand.

Mostert, M. (1993) *n Kwantitatiewe en Kwalitatiewe Waardebepaling van Ondernemingsrisiko en -mislukking.* D.Com: Randse Afrikaanse Universiteit.

Neophytou, E. & Molinero, C.M. (2001) "Predicting Corporate Failure in the UK: A Multidimensional Scaling Approach". *Discussion Papers in Accounting and Management Science No 01-172.* University of Southampton, England.

Payne, T.R., Edwards, P. & Green, C.L. (1995) "Experience with Rule Induction and k-Nearest Neighbour Methods for Interface Agents that Learn". *Working Notes of the ICML'95 Workshop on "Agents that learn from other agents".*

Rifkin, R.M. (2002) *Everything Old is New Again: A fresh look at historical approaches to machine learning.* Ph.D. Thesis, Massachusetts Institute of Technology.

Sutton, J. (1995) *The Relationship between Business Failure and Corporate Governance in South Africa.* MBA: University of Cape Town.

Truter, W. (1996) *Forecasting Corporate Failure using Financial Ratios: A Z-Score Calculation for Non-Listed Companies in South Africa.* MBA: University of Cape Town.

Van Eyden, R.J. (1994) *The Application of Neural Networks in the Forecasting of Share Prices.* D.Com: University of Pretoria.

Walters, P. (1982) *Management Experience as a Factor in Corporate Failure among Small Retailers.* MBA: University of Pretoria.

Conference Proceedings

Baluja, S. & Caruana, R. (1995) "Removing the Genetics from the Standard Genetic Algorithm". In A. Prieditis and S. Russel ed., International Conference on Machine Learning, San Mateo, California, Morgan Kaufmann Publishers, 38-46.

Cantú-Paz, E. (2002) "Feature subset selection by estimation of distribution algorithms." In *GECCO-2002: Proceedings of the Genetic and Evolutionary Computation Conference*, (Eds) W. B. Langdon, E. Cantu-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke & N. Jonoska, San Francisco, Morgan Kaufmann, 754.

FrieB, T. & Harrison, R.F. (1999) "A Kernel-Based Adaline". *ESANN'1999 Proceedings – European Symposium on Artificial Neural Networks:* Bruges (Belgium), 21-23 April, D-Facto Public, 245-250.

Fung, G. & Mangasarian, O. (2001) "Proximal support vector machine classifiers". In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 2001, 786.

Greene, J.R. (1996) "Population-Based Incremental Learning as a Simple Versatile Tool for Engineering Optimisation". In *Proceedings of the First International Conference on Evolutionary Computation and its Applications (EvCA '96)*, 258-269.

Greene, J.R. (2001a) "Feature Subset Selection using Thornton's Separability Index and its Applicability to a Number of Sparse Proximity-Based Classifiers". In *Proceedings of the Twelfth Annual Symposium of the Pattern Recognition Association of South Africa*, Ed. L. Botha, 29-30 November 2001, 1-10.

John, G., Kohavi, R. & Phleger, K. (1994) "Irrelevant features and the feature subset problem." In *Proceedings of the 11th International Conference on Machine Learning,* Morgan Kaufmann, 121-129.

Odom, M. & Sharda, R. (1990) "A Neural Network Model for Bankruptcy Prediction". In *Proceedings of the IEEE International Conference on Neural Networks II*, 163-168.

Piramuthu, S. (1999) "The Hausdorff Distance Measure for Feature Selection in Learning Applications". In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999, 1-6.

Schölkopf, B., Herbrich, R. & Smola, A.J. (2001) "A Generalised Representer Theorem". In *Proceedings of the 14th Annual Conference on Computational Learning Theory* 416-426.

Thornton, C. (1997) "Separability is a learner's best friend". In J.A. Bullinaria, D.W. Glasspool, and G. Houghton, editors, *Proceedings of the Fourth Neural Computation and Psychology Workshop: Connectionist Representations*, 40-47. Springer-Verlag.

Websites

www.moodysqra.com Moody's Quantitative Risk's Public Firm Risk Model.

Sarle, W.S. (2000) "Neural Networks FAQ". ftp://ftp.sas.com/pub/neural/FAQ.html.

Olmsted, D.D. (1998) *History and Principles of Neural Networks.* www.neurocomputing.org/history.htm.

APPENDICES

APPENDIX A - MATLAB CODE

A.1. MATLAB CODE: THORNTON'S SEPARABILITY INDEX

Source: Adapted from Greene (2001a, 2)

function [s] = sepindx(X, t)

```
% SEPINDX  Thornton's Separability Index
%       This function calculates the proportion of points in the input matrix that have a nearest
%       neighbour with the same input label.

% [sepindex] = sepindx ( Input Matrix, Input Labels )

% Accepts a matrix X in which each row is a vector of numeric features (usually normalised
% into the range 0-1)

% t is a vector of labels (usually +1 or -1)

% Returns s, a number between 0 and 1, which is a measure of the degree to which the
% classes are geometrically separable. s is simply the fraction of instances whose
% classification label is shared by its nearest-neighbour (determined on the basis of simple
% Euclidean distance)

D2 = dist2(X,X);                      % calculates Euclidean distance between all points

[D2Sort, Idx] = sort(D2);             % sort D2

s = sum( t(Idx(1,:)) == t(Idx(2,:)) )...
   / length(t);                       % fraction of points with same class n.neighbour
```

A.2. MATLAB CODE: POPULATION-BASED INCREMENTAL LEARNING FOR FEATURE SUBSET SELECTION USING SEPARABILITY INDEX AS EVALUATION CRITERION

Source: Adapted from Greene (2001) and Baluja & Caruana (1995, 7)

```
function [ bestmask, bestsi, rec ] = pbilmask ( X, t, epoch, ntrials )

% PBILMASK  This function calculates the best input variable mask so as to maximise the
%           separability index between the input points.

% [bestmask, sepindex, rec ] = pbilmask (input matrix, input labels, number of generations,
% number of trials)

% Default epochs are 50 and default trials per epoch are 10.

[nsets, nvars] = size(X);        % measures the number of rows and columns
bestsi = -inf;                   % initialise the best sep index
rec = [];                        % initialise a matrix to record improvement in the sep index
pv = 0.5*ones(1, nvars);         % initialise probability vector
s = sepindx(X,t);                % initialise separability index
if nargin < 3, epoch = 50;, end  % set epoch to default
if nargin < 4, ntrials = 10;, end % set ntrials to default

for gen = 1:epoch
  for trial = 1:ntrials
    mask = rand(1, nvars) < pv;% sample pv to form mask
    XX = X*diag(mask);           % apply mask to input
    s = sepindx(XX,t);           % evaluate this subset
    if s>bestsi                  % if its an improvement
      bestsi = s;                % record it as best
      bestmask = mask;           % and store the mask
    end
  end
  rec=[rec, bestsi];             % record the improvement in sepindx
  pv = 0.9*pv + 0.1*bestmask;    % and adapt the pv
  pv = pv - (pv-0.5)./200;       % and maintain diversity
end
plot(rec)                        % plot the improvement in sepindx against epochs
```

A.3. MATLAB CODE: *K*-NEAREST NEIGHBOUR MODEL CONSTRUCTION

```
function m = testXfoldknn (X, t, C, startk, endk)

% TESTXFOLDKNN  This function tests the error on the k-nearest neighbour classifier using
%               using X-fold cross-validation for a series of values for k and returns m, a
%               matrix with the knn accuracy for each value of k tested.
%
% function m = testXfoldknn (X, t, C, startk, endk)
%               X = input matrix
%               t = target matrix
%               C = number of folds for cross-validation
%               startk = lowest odd number to use as k in the knn classifier (default = 1)
%               endk   = highest odd number to use as k in the knn (default = 21)
%
% The knn classifier is run on the data for all odd incremental values of k between startk and
% endk.

if nargin < 5; endk = 21; end       % default endk = 21
if nargin < 4; startk = 1; end      % default startk = 1
if nargin < 3; C = 10; end          % default set to tenfold cross-validation

m = [];                             % initialise m
k = [];                             % initialise k

for k = startk:2:endk               % for all odd value of k between startk and endk
    e = Xfoldknn (X, t, k, C);      % run knn classifier counting errors using Xfoldcross (see
                                    % code below)
    m = [m; k e ];                  % append k and related error onto m
end

kt = m (:, 1);                      % separate k and e
et = m (:, 2);
plot (kt, et)                       % plot e against k
```

```
function err = Xfoldknn (X, t, k, C)

% XCROSSFOLD        This function uses X-fold cross-validation on a k-nearest neighbour
%               classifier to calculate the accuracy of the k-NN model
%
% function e = Xfoldknn (X, t, k, C)
%               X = input matrix
%               t = target matrix
%               k = number of nearest neighbours to use in classifier
%               C = number folds for cross-validation
%
%               err = percentage of errors

if nargin < 4; C = 10; end          % default 10-fold cross validation
if nargin < 3; k = 1; end           % default set to nearest neighbour classifier
D = [X, t];                         % assemble data matrix
[p,d] = size(D);                    % get total number of instances and columns
ptr = fix((1-(1/C))*p);             % no. of training instances (90%)
pte = p - ptr;                      % no. of test instances (the rest, 10%)
e = 0;                              % initialise error count to zero

for split = 1:C                     % for each of the 10 splits
        Dtr = D(1:ptr, : );         % split off training data
        Dte = D(ptr+1:p,:);         % split off test data
```

```
        Xtr = Dtr( : , 1: d-1);       % input data: first d-1 columns
        ttr = Dtr( : , d);            % target data last (dth column)
        Xte = Dte (: , 1: d-1);       % input data
        tte = Dte( : ,d);             % target data

        y = knn(k, Xtr, ttr, Xte); % train the classifier (see code below); predict the test data
        e = e + sum(y ~= tte);    % count the errors and accumulate them

        D = [Dte;Dtr];                % reassemble data matrix with test data at top
end

err = e / p * 100;                    % calculate the percentage error
```

- -

```
function [err, Y, T] = XfoldknnYT (X, t, k, C)

% XFOLDKNNYT        This function uses X-fold cross-validation on a k-nearest neighbour
%                   classifier to calculate the accuracy of the k-NN model. In addition it collects
%                   all predicted classifications as determined by the kNN model
%
% function [err, Y, T] = XfoldknnYT (X, t, k, C)
%                   X = input matrix
%                   t = target matrix
%                   k = number of nearest neighbours to use in classifier
%                   C = number folds for cross-validation
%
%                   err = percentage of errors
%                   Y = the predicted classification of the test data
%                   T = the true classification of the test data that corresponds to Y

if nargin < 4; C = 10; end            % default 10-fold cross validation
if nargin < 3; k = 1; end             % default set to nearest neighbour classifier

D = [X, t];                           % assemble data matrix
[p,d] = size(D);                      % get total number of instances and columns
ptr = fix((1-(1/C))*p);               % no. of training instances (90%)
pte = p - ptr;                        % no. of test instances (the rest, 10%)
e = 0;                                % initialise error count to zero
Y = [];                               % initialise Y
T = [];                               % initialise T

for split = 1:C                       % for each of the 10 splits
        Dtr = D(1:ptr, : );           % split off training data
        Dte = D(ptr+1:p,:);           % split off test data

        Xtr = Dtr( : , 1: d-1);       % input data: first d-1 columns
        ttr = Dtr( : , d);            % target data last (dth column)
        Xte = Dte (: , 1: d-1);       % input data
        tte = Dte( : ,d);             % target data
        T = [T; tte];                 % collect target data true classifications on each epoch

        y = knn(k, Xtr, ttr, Xte); % train the classifier (see code below); predict the test data
        Y = [Y; y];                   % collect predicted classification of target data

        e = e + sum(y ~= tte);        % count the errors and accumulate them

        D = [Dte;Dtr];                % reassemble data matrix with test data at top
end

err = e / p * 100;                    % calculate the percentage error
```

```
function y = knn(k, X, t, Q)

% KNN        This function predicts the classification of an item based on an average of its
%            k-nearest neighbours

% function y = knn(k, X, t, Q)
%            k = number of nearest neighbours to be considered (recommend odd number
%            1..27 )
%            X = input data (row: instance; column: feature. Normalised column-wise
%            (mean 0 , std 1))
%            t = labels  (can be any discrete values; typically -1 and +1)
%            Q = matrix of query points (must have same number of columns as X)
%
%            y:  predicted labels for instances in Q

D2 = dist2(X,Q);                 % col i: sq. dist from ith query point to each training point
[r, c] = size (Q);


[S,I] = sort(D2);                % S has nearest-neighbor distances on top row
I = I(1:k,:)';                   % select kth top rows of index matrix; transpose into columns
y = t(I(:,1:k));                 % look up labels of k nearest training points to each query
                                 % point.

if k > 1
   if r == 1
      y =sign(sum(y)...          % majority vote
       + 0.1*randn(size(1,1),1)); % sign(rowsum + tie-breaker)
   else
      y =sign(sum(y,2)...         % majority vote
      + 0.1*randn(size(y,1),1));  % sign(rowsum + tie-breaker)
   end
end
```

A.4. MATLAB CODE: KERNEL RIDGE REGRESSION MODEL CONSTRUCTION

```
function m = testsigmaXfoldkrr (X, t, gamma, C, startsigma, incrsigma, endsigma)

% TESTSIGMAXFOLDKRR     This function performs C-fold cross-validation using Kernel
%                Ridge Regression for a series of values for sigma.
%
% function m = testsigmaXfoldkrr (X, t, gamma, C, startsigma, incrsigma, endsigma)
%                X              = Input Matrix (Normalised)
%                t              = Target Matrix (+/-1)
%                gamma          = Regularisation Parameter
%                C              = Number of folds for cross-validation
%                startsigma     = The lowest value for sigma at which to start the search
%                incrsigma      = The incremental value added to each new sigma tested
%                endsigma       = The highest value for sigma at which to end the search
%
% The krr classifier is run on the data for all incremental values of sigma between startsigma
% and endsigma in increments of incrsigma.

[p, d] = size (X);                          % determine dimensions of matrix X

if nargin < 7; endsigma = 2; end            % default end point set to sigma = 1
if nargin < 6; incrsigma = 0.05; end        % default increments set to 0.05
if nargin < 5; startsigma = 0.05; end       % default start point set to sigma = 0.05
if nargin < 4; C = p; end                   % default set to LOO validation
if nargin < 3; gamma = 1; end               % default gamma set to 1

m = [];                                     % initialise m

for sigma = startsigma: incrsigma: endsigma     % for all values of sigma increasing at ...
                                                % incrsigma btwn startsigma and endsigma
        [err, Y, T] = XfoldkrrYT (X, t, sigma,...  % run krr classifier and count percentage ...
                gamma, C);                      % errors based on C-fold cross-validation
        m = [m; sigma err];

end

plot (m(:,1), m(:,2));                      % plot percentage errors against sigmas used
```

```
function m = testgammaexp (X, t, sigma, C, startg, endg)

% TESTGAMMAEXP   This function performs C-fold cross-validation using Kernel Ridge
%                Regression for a series of values for gamma of the order 10ex.
%
% function m = testgammaexp (X, t, sigma, C, startg, endg)
%                X        = Input Matrix (Normalised)
%                t        = Target Matrix (+/-1)
%                sigma    = Kernel Width
%                C        = Number of folds for cross-validation
%                startg   = The lowest value for exp at which to start the search
%                endg     = The highest value for exp at which to end the search
%
% The krr classifier is run on the data for all incremental values of gamma between
% startgamma and endgamma in increments of 10ex.

[p, d] = size (X);                          % determine the dimensions of matrix X

if nargin < 6; endg = 6; end                % default end point set to exp = 6
if nargin < 5; startg = -6; end             % default start point set to exp = -6
```

```
if nargin < 4; C = p; end                    % default set to LOO validation
if nargin < 3; sigma = 0.5; end              % default sigma set to 0.5

m = [];                                      % initialise m
x = [startg:1:endg];
z = 10.*(ones(1,length(x)));
g = z.^x;

for gamma = g                                % for all values of gamma increasing at incrgamma
                                             % between startgamma and endgamma
        [err, Y, T] = XfoldkrrYT (X, t, sigma,...    % run krr classifier and count % errors ...
                gamma, C);                           % based on C-fold cross-validation
        m = [m; gamma err];

end

plot (m(:,1), m(:,2));                       % plot percentage errors against gammas used
```

---

function m = testgammaXfoldkrr (X, t, sigma, C, startgamma, incrgamma, endgamma)

```
% TESTGAMMAXFOLDKRR    This function performs C-fold cross-validation using Kernel
%                Ridge Regression for a series of values for gamma.
%
% function m = testgammaXfoldkrr (X, t, sigma, C, startgamma, incrgamma, endgamma)
%                X               = Input Matrix (Normalised)
%                t               = Target Matrix (+/-1)
%                sigma           = Kernel Width
%                C               = Number of folds for cross-validation
%                startgamma      = The lowest value for gamma at which to start the search
%                incrgamma       = The incremental value added to each new gamma tested
%                endgamma        = The highest value for gamma at which to end the search
%
% The krr classifier is run on the data for all incremental values of gamma between
% startgamma and endgamma in increments of incrgamma.

if nargin < 7; endgamma = 2; end             % default end point set to gamma = 2
if nargin < 6; incrgamma = 0.2; end          % default increments set to 0.2
if nargin < 5; startgamma = 0; end           % default start point set to gamma = 0
if nargin < 4; C = 10; end                   % default set to tenfold cross-validation
if nargin < 3; sigma = 0.5; end              % default sigma set to 0.5

m = [];                                      % initialise m

for gamma = startgamma:incrgamma:endgamma            % for all values of gamma ...
                                                     % increasing at incrgamma btwn ...
                                                     % startgamma & endgamma
        [err, Y, T] = XfoldkrrYT (X, t, sigma,...    % run krr classifier and count % errors ...
                gamma, C);                           % based on C-fold cross-validation
        m = [m; gamma err];

end

plot (m(:,1), m(:,2));                               % plot % errors against gammas used
```

---

function [err, Y, T] = XfoldkrrYT (X, t, sigma, gamma, C)

```
% XFOLDKRRYT       This function performs C-fold cross-validation on D using Kernel
%                Ridge Regression
```

```
%
% function [err, Y, T] = XfoldkrrYT (X, t, sigma, gamma, C)
%               X       = Input Matrix (Normalised)
%               t       = Target Matrix (+/-1)
%               sigma   = Kernel Width (default = 0.5)
%               gamma   = Regularisation Parameter (default = 1)
%               C       = Number of folds for cross-validation (default = LOO validation)
%
%               err     = Percentage of errors
%               Y       = Predicted class of data point using krr
%               T       = True class of data point

D = [X, t];                         % assemble data matrix
[p,d] = size(D);                    % get total number of instances and columns

if nargin < 5; C = p; end           % default LOO cross-validation
if nargin < 4; gamma = 1; end       % default gamma = 1
if nargin < 3; sigma = 0.5; end     % default sigma = 0.5

ptr = fix((1-(1/C))*p);             % no. of training instances
pte = p - ptr;                      % no. of test instances (the rest)
e = 0;                              % initialise error count to zero
Y = [];                             % initialise Y
T = [];                             % initialise T

for datasplit = 1:C                 % for each of the C data-splits

        Dtr = D(1:ptr, : );         % split off training data
        Dte = D(ptr+1:p,:);         % split off test data

        Xtr = Dtr( : , 1: d-1);     % input data: first d-1 columns
        ttr = Dtr( : , d);          % target data last (d^th column)
        Xte = Dte (: , 1: d-1);     % input data
        tte = Dte( : ,d);           % target data
        T = [T; tte];               % collect target data true classifications on each epoch

        y = krr (Xtr, ttr, Xte, sigma,...
                gamma);             % train the krr classifier and predict test data class

        Y = [Y, y];                 % collect predicted classification of target data

        e = e + sum(y ~= tte);      % count the errors and accumulate them

        D = [Dte;Dtr];              % reassemble data matrix with test data at top

end

err = e / p * 100;                  % calculate the percentage error
```

- -

Source: Greene (2003)

function y = krr (Xtr, ttr, Xte, sigma, gamma)

```
% KERNEL RIDGE REGRESSION            y = krr (Xtr, ttr, Xte, sigma, gamma)
%
%     input:                         training data: Xtr, ttr (X normalised, t = +/-1)
%     test data:                     Xte (X normalised, t = +/-1)
%     kernel width:                  sigma
%     regularisation parameter:      gamma
%
%     output: predicted Xte class:   y

% TRAIN (find regression weights alpha)
%     Finds the weighted sum of training-point kernels to force a fit to the training data
%     [ Xtr, ttr]

D2 = dist2 (Xtr, Xtr);               % training-point pair-wise distances
K = exp (-D2 / (2*sigma^2));         % kernel activation values at training points
K = K + gamma*eye(size (K));         % regularisation
alpha = K\ttr;                       % solve for regression weights

% TEST  (predict test targets as y)
%     Evaluates this weighted sum at each test point; uses the sign of the function as a
%     predictor of the test point classification

D2 = dist2 (Xte, Xtr);               % pair-wise training/test point distances
K = exp (-D2 / (2*sigma^2));         % kernel activations at test points
y = sign (K*alpha);                  % sign of summed weighted kernel activations
```

APPENDIX B – LITERATURE REVIEW SUMMARY

This appendix consists of the following two parts:

- **Appendix B.1**: Summary table of key corporate failure studies.
- **Appendix B.2**: Summary table of ratios used in key corporate failure studies.

Appendix B.1. Summary table of key corporate failure studies.

#	Author	Year	Country	Definition of Failure	Purpose	Sample Period	Industry Type	Method	Classification Type	Sample Size	Sample Selection	Predictor Variables
1	Beaver	1966	U.S.	Inability of firm to pay its debts as they mature	Test predictive ability of financial ratios	1954-1964	Industrial Listed	Univariate Analysis	Dichotomous	79 Failed 79 Non-F	Paired sample method	Firm-specific financial
2	Beaver	1968	U.S.	Inability of firm to pay its debts as they mature	Test predictive ability of stock price changes	1954-1964	Industrial Listed	Univariate Analysis	Dichotomous	79 Failed 79 Non-F	Paired sample method	Firm-specific stock price
3	Altman	1968	U.S.	Chapter X Bankruptcy	Assess the quality of ratio analysis as an analytical tool	1946-1965	Manufacturing	MDA: linear	Dichotomous	33 Failed 33 Non-F	Paired sample method	Firm-specific financial
4	Deakin, E.B.	1972	U.S.	Chapter X Bankruptcy, Insolvency, Liquidated for the benefit of creditors	Alternative model to Beaver and Altman	1964-1970	Not stated	MDA: linear	Dichotomous	32 Failed 32 Non-F	Non-paired sample method	Firm-specific financial, Stock Returns
5	Blum, M.	1974	U.S.	Chapter X Bankruptcy, agreement with creditors to reduce debt, insolvency	Develop a Failing Company Model to aid U.S. antitrust investigations	1954-1968	Industrial Listed	MDA: linear	Dichotomous	115 Failed 115 Non-F	Paired sample method	Firm-specific financial
6	Libby, R.	1975	U.S.	Chapter X Bankruptcy, Insolvency, Liquidated for the benefit of creditors	See if accounting ratios provide useful information to loan officers	1964-1970	Not stated	behavioural, loan officers	Dichotomous	30 Failed 30 Non-F	Non-paired sample method	Firm-specific financial
7	Elam, R.	1975	U.S.	Chapter X or XI Bankruptcy, Agreement with creditors to reduce debt	Determine if lease capitalisation enhances accuracy of failure prediction model	1966-1972	Industrial, Listed	MDA: linear	Dichotomous	48 Failed 48 Non-F	Paired sample method	Firm-specific financial
8	Altman, E.I., Haldeman, R.G. & Narayanan, P.	1977	U.S.	Chapter X Bankruptcy	New model that explicitly considers recent developments	1969-1975	Manufacturing, Retailing	MDA: linear & quadratic	Dichotomous	53 Failed 58 Non-F	Paired sample method	Firm-specific financial
9	Norton, C.L. & Smith, R.E.	1979	U.S.	Chapter X or XI Bankruptcy	To compare the prediction of bankruptcy based the general price level and historical cost financial statements	1971-1975	Industrial, Listed	MDA: linear	Dichotomous	30 Failed 30 Non-F	Paired sample method	Firm-specific financial
10	Norton, C.L. & Smith, R.E.	1979	U.S.	Chapter X or XI Bankruptcy	To compare the prediction of bankruptcy based the general price level and historical cost financial statements	1971-1975	Industrial, Listed	MDA: linear	Dichotomous	30 Failed 30 Non-F	Paired sample method	Firm-specific financial and inflation
11	Sharma, S. & Mahajan, V.	1980	U.S.	Chapter X or XI Bankruptcy	To illustrate a [illegible] failure model	1970-1976	Retailing	MDA: linear	Dichotomous	23 Failed 23 Non-F	Paired sample method	Firm-specific financial
12	Ohlson, J.A.	1980	U.S.	Chapter X or XI Bankruptcy	To construct a logit prediction model using unmatched data set	1970-1976	Industrial, Listed	Logit Model	Dichotomous, Probability of failure	105 Failed 2058 Non-F	Non-paired sample method	Firm-specific financial
13	Whittred, G. & Zimmer, I.	1984	Australia	Not given	Use of accounting [illegible] to predict failure of Australian firms	1964-1978	Industrial, Listed	MDA: linear	Dichotomous	32 Failed 32 Non-F	Paired sample method	Firm-specific financial and non-financial
14	Frydman, H., Altman, E.I. & Kao, D.	1985	U.S.	Chapter X or XI Bankruptcy	Presents a new classification procedure, RPA and compares to MDA	1971-1981	Manufacturing, Retailing	RPA (3)	Dichotomous	58 Failed 142 Non-F	Non-paired sample method	Firm-specific financial
15	Frydman, H., Altman, E.I. & Kao, D.	1985	U.S.	Chapter X or XI Bankruptcy	Presents a new classification procedure, RPA and compares to MDA	1971-1981	Manufacturing, Retailing	RPA (2)	Dichotomous	58 Failed 142 Non-F	Non-paired sample method	Firm-specific financial
16	Frydman, H., Altman, E.I. & Kao, D.	1985	U.S.	Chapter X or XI Bankruptcy	Presents a new classification procedure, RPA and compares to MDA	1971-1981	Manufacturing, Retailing	MDA: linear	Dichotomous	58 Failed 142 Non-F	Non-paired sample method	Firm-specific financial
17	Zavgren, C.V.	1985	U.S.	Chapter X or XI Bankruptcy	To use the logit technique to test a new bankruptcy model and develop a methodology for evaluating the significance of probability of [illegible]	1972-1978	Manufacturing	Logit Model	Dichotomous	45 Failed 45 Non-F	Paired sample method	Firm-specific financial
18	Lane, W.R., Looney, S.W. & Wansley, J.W.	1986	U.S.	Payoff of depositors	Apply the cox proportional hazards model to bank failure prediction	1979-1983	Banks	Cox Prop. Hazards	Dichotomous, Time to failure	130 Failed 334 Non-F	Paired sample method	Firm-specific financial
19	Peel, M.J., Peel D.A. & Pope, P.F.	1986	U.K.	Voluntary liquidations by creditors, Receivership, Winding-up by the court	Extended results to the U.K. combining firm-specific non-financial [illegible] from company annual report	1974-1982	Manufacturing	Logit Model	Dichotomous	34 Failed 44 Non-F	Non-paired sample method	Firm-specific financial and non-financial
20	Lau, A.H.	1987	U.S.	Classified in five-state continuum of distress	Model that calculates the probability of a firm falling into one of five states on a continuum of corporate financial health	1971-1977	General	Multinomial Logit Model	Five-state, probability of failure	350 Non-F 50 Financial Distress	Non-paired sample method	Firm-specific financial, industry, [illegible]
21	Looney, S.W., Wansley, J.W. & Lane, W.R.	1989	U.S.	Payoff of depositors	[illegible] banks with higher exposure to business failure	1984-1986	Banks	Cox, MDA, Logit and quadratic	Dichotomous, Time to failure	289 Failed 377 Non-F	Non-paired sample method	Firm-specific financial
22	Platt, H.D., Platt, M.B. & Pedersen, J.G.	1994	U.S.	Chapter X Bankruptcy	A multi-year sample is tested for [illegible] oil and gas firms using alternate model specifications of deflators and real [illegible]	1982-1988	Oil & Gas	Logit Model	Dichotomous	35 Failed 89 Non-F	Non-paired sample method	Firm-specific financial
23	Platt, H.D., Platt, M.B. & Pedersen, J.G.	1994	U.S.	Chapter X Bankruptcy	[illegible] oil and gas firms using alternate model specifications of deflators and real [illegible]	1982-1988	Oil & Gas	Logit Model	Dichotomous	35 Failed 89 Non-F	Non-paired sample method	Firm-specific financial and deflators
24	Platt, H.D., Platt, M.B. & Pedersen, J.G.	1994	U.S.	Chapter X Bankruptcy	A multi-year sample is tested for [illegible] oil and gas firms using alternate model specifications of deflators and real [illegible]	1982-1988	Oil & Gas	Logit Model	Dichotomous	35 Failed 89 Non-F	Non-paired sample method	Firm-specific financial and [illegible]
25	Platt, H.D., Platt, M.B. & Pedersen, J.G.	1994	U.S.	Chapter X Bankruptcy	[illegible] oil and gas firms using alternate model specifications of deflators and real [illegible]	1982-1988	Oil & Gas	Logit Model	Dichotomous	35 Failed 89 Non-F	Non-paired sample method	Firm-specific financial, deflators, and [illegible]
26	Boritz, J.E. & Kennedy, D.B.	1995	U.S.	Chapter X or XI Bankruptcy	Examines the effectiveness of different neural networks [illegible] MDA and logit [illegible] in predicting bankruptcy	1971-1984	Not given	Neural Network	Dichotomous	171 Failed 6153 Non-F	Non-paired sample method	Firm-specific financial
27	Lennox, C.	1999	U.K.	Liquidated, Receivership	Evaluates well-specified logit/probit models over MDA, & analyses the effects of industry sector and business cycle on probability of bankruptcy	1987-1994	Listed	Logit, Probit, MDA	Dichotomous	90 Failed 949 Non-F	Non-paired sample method	Firm-specific financial
28	Atiya, A.F.	2001	U.S.	Chapter X or XI Bankruptcy	Compare a neural network model using financial ratios and equity-based system to a model using financial ratios alone	Not given	Not given	Neural Network	Dichotomous	195 Failed 716 Non-F	Non-paired sample method	Firm-specific financial
29	Atiya, A.F.	2001	U.S.	Chapter X or XI Bankruptcy	Compare a neural network model using a financial ratio and equity-based system to a model using financial ratios alone	Not given	Not given	Neural Network	Dichotomous	195 Failed 716 Non-F	Non-paired sample method	Firm-specific financial, Stock price data

Appendix B.2. Summary table of ratios used in key corporate failure studies.

#	Author	WC/TA	WC/TD	TD/TA	FD>/TA	CF/TD	CF/TA	NI/TA	NI/TD	NI/BV	CA/S	CA/CL	QA/CL	QA/CA	CAS/CL	CAS/S	QE/S	CAS/TA	EBI/TA	EBIT/TA	RE/TA	S/TA	S/FA	S/TD
1	Beaver	✓		✓		✓		✓				✓												
2	Beaver																							
3	Altman	✓																	✓		✓	✓		
4	Deakin, E.B	✓		✓		✓		✓				✓	✓		✓	✓		✓						
5	Blum, M.					✓														✓				
6	Libby, R.						✓	✓			✓	✓												
7	Elam, R.			✓		✓	✓		✓	✓		✓	✓		✓	✓						✓		
8	Altman, E.I., Haldeman, R.G & Narayanan, R.											✓							✓	✓	✓			
9	Norton, C.L. & Smith, R.E.			✓						✓	✓					✓								
10	Norton, C.L. & Smith, R.E.					✓					✓					✓								
11	Sharma, A & Mahajan, V.											✓							✓					
12	Ohlson, J.A.	✓		✓	✓			✓				✓			✓									
13	Whittred, G. & Zimmer, I.					✓		✓				✓	✓						✓					
14	Frydman, H., Altman, E.I. & Kao, D.			✓		✓										✓								
15	Frydman, H., Altman, E.I. & Kao, D.			✓		✓										✓								
16	Frydman, H., Altman, E.I. & Kao, D.							✓				✓	✓						✓					
17	Zavgren, C.V.									✓				✓				✓					✓	
18	Lane, W.R., Looney, S.W. & Wansley, J.W.																							
19	Peel, M.J., Peel, D.A. & Pope, P.F.	✓				✓																		✓
20	Lau, A. H.		✓														✓							
21	Looney, S.W., Wansley, J.W. & Lane, W.R.																							
22	Platt, H.D., Platt, M.B. & Pedersen, J.G.			✓			✓																	
23	Platt, H.D., Platt, M.B. & Pedersen, J.G.			✓			✓																	
24	Platt, H.D., Platt, M.B. & Pedersen, J.G.			✓			✓																	
25	Platt, H.D., Platt, M.B. & Pedersen, J.G.			✓			✓																	
26	Bontz, J.C. & Kennedy, D.B.	✓		✓	✓			✓				✓			✓				✓		✓	✓		
27	Lennox, C.			✓			✓	✓							✓									
28	Atiya, A.F.						✓	✓											✓					
29	Atiya, A.F.						✓	✓											✓					
	Total	6	1	13	2	9	9	10	1	3	3	10	4	1	5	6	1	2	8	2	3	3	1	1

Appendix B.2. continued...

#	Author	S/DR	QA/TA	QA/I	QB/I	CA/TA	CL/TA	CL/TD	MVE/TD	UVL/TD	INT/TA	BVE/MVE	EBI/IE	MVL/TD	LOG (TA)	GNP/TA	CASH FLOW	SHAR ROR	STD (NI)	OTHER	NON-LINEAR	DEFLATOR	MACRO
1	Beaver																						
2	Beaver																						
3	Altman								✓														
4	Deakin, E.B.		✓			✓																	
5	Blum, M.			✓					✓	✓							✓	✓		✓			
6	Libby, R.					✓																	
7	Bam, R.						✓			✓													
8	Altman, E.I., Haldeman, R.G. & Narayanan, R.												✓	✓	✓				✓				
9	Norton, C.L. & Smith, R.E.	✓				✓				✓										✓			
10	Norton, C.L. & Smith, R.E.					✓				✓										✓		✓	
11	Sharma, A & Mahajan, V.																						
12	Ohlson, J.A.														✓	✓			✓	✓			
13	Whittred, G. & Zimmer, I.								✓	✓		✓	✓										
14	Frydman, H., Altman, E.I. & Kao, D.		✓										✓	✓									
15	Frydman, H., Altman, E.I. & Kao, D.		✓										✓	✓									
16	Frydman, H., Altman, E.I. & Kao, D.		✓			✓							✓	✓	✓								
17	Zavgren, C.V.				✓					✓													
18	Lane, W.R., Looney, S.W. & Wansley, J.W.																						
19	Peel, M.J., Peel, D.A. & Pope, P.F.															✓					✓		
20	Lau, A. H.									✓								✓		✓			
21	Looney, S.W. Wansley, J.W. & Lane, W.R.																						
22	Platt, H.D., Platt, M.B. & Pederson, J.G.							✓												✓			
23	Platt, H.D., Platt, M.B. & Pederson, J.G.							✓												✓		✓	
24	Platt, H.D., Platt, M.B. & Pederson, J.G.							✓												✓			✓
25	Platt, H.D., Platt, M.B. & Pederson, J.G.							✓												✓		✓	✓
26	Boritz, J.E. & Kennedy, D.B.								✓						✓	✓			✓	✓			
27	Lennox, C.	✓																		✓			
28	Atiya, A.F.										✓						✓						
29	Atiya, A.F.										✓						✓	✓					
	Total	2	4	1	1	5	1	4	4	7	2	1	5	4	4	3	3	3	3	11	1	3	2

APPENDIX C – BFA ANNUAL FINANCIAL STATEMENT TEMPLATE

BFA RAID STATION STANDARD FINANCIAL STATEMENT DATABASE TEMPLATE

BALANCE SHEET ('000s)

1 Ord Shareholders Interest

2 Ord Share Capital
3 Share Premium
4 Non-distrib Reserves
5 Distrib Reserves

6 Adj Mkt/dir Val In Inv.

7 Ord Sharehldrs Aft Adj.
8 Preference Shares

9 Irredeemable
10 Redeemable
11 Convertible

12 Outside Shareholders Int.
13 Total Shareholders Int.
14 Deferred Tax
15 Other
16 Long Term Liabilities

17 Convertible Debentures
18 Dir/sharehldrs Loans
19 Non Interest Bearing
20 Interest Bearing

21 Capital Employed
22 Total Liabilities

23 Fixed Assets
24 Mining Assets
25 Intangible Assets

26 Goodwill
27 Patents & Trademarks
28 Cost Of Control
29 Other

30 Non Current Assets
31 Investments & Loans

32 Inv At Cost/Market Val
33 Long Term Loans

34 Current Assets

35 Inventory
36 Debtors
37 Cash & Near Cash
38 Dividends
39 Tax
40 Other

41 Current Liabilities

42 Creditors
43 Dividends
44 Tax
45 Interest Bearing
46 Non Interest Bearing

47 Net Current Assets
48 Adj Mkt/dir Val In Inv.
49 Employment Of Capital
50 Total Assets

GENERAL SUPPLEMENTARY ('000s)

201 Shares In Issue Y/E Ord
202 Shares In Issue Y/E 'N'
203 Shares In Issue Y/E 'A'
204 Shares In Issue Y/E 'B'
206 Shares In Issue Wgt Ave
207 Shares In Issue Ful Dil
208 Revaluation Reserve
209 Minority Reval Reserve
210 Minority Equity Acc Res
211 Commitments - Land Bldg
212 Commitments - Other
213 Foreign Borrowings
214 Convertible Pref Shares
215 Convertible Deb & Loans
216 Share In Issue Latest
217 Mining Assets at Cost
218 Depn/Amort Mine Assts
219 Medical Aid Liabs
220 Pension Fund Liabs
221 LT Loans - Int Bear
222 LT Loans - Int Free
223 ST Loans - Int Bear
224 ST Loans - Int Free
225 Asst Reval Surp Cur Yr
226 Pft/Loss Forex Translate
227 Pft/Loss Forex Transact

INCOME STATEMENT ('000s)

60 Turnover
61 % Change In Turnover

62 Investment Income
63 Operating Profit
64 Interest Received

65 Gross Income
66 Interest & Fin Chngs(-ve)

67 Taxation (-ve)

68 Current (-ve)
69 Deferred (-ve)
70 Other (-ve)

71 Profit After Int & Tax
72 Pref. Dividends (-ve)
73 Minority Interest (-ve)
74 Associate Companies
75 Discontinued Operations
76 Other
77 Convertible Deb Int (-ve)

78 Profit Attrib To Ord Shrs
79 Extra Ordinary Items
80 Bottom Line Earnings
81 Dividends Paid (-ve)

82 Ret. Earnings-current Yr
83 Trans To Res/life Funds
84 Ret. Earnings-prior Year

85 Total Reserve

86 H/L Earngs Per Shr
87 Dividends Per Shr
88 Depreciation
89 Audit Fees
90 Directors Emoluments

GENERAL SUPPLEMENTARY ('000s)

301 Lease Chrge - Land Bldg
302 Lease Chrge - Other
303 Research & Development
304 EPS-Equit Accnt
305 EPS-Bottom Line
306 EPS-Headline
307 EPS-Fully Dil H/L
308 EPS-Fully Dil B/L
309 Effect Tax Rate
310 Deferred Tax - Cont Lib
311 Deferred Tax - Current
312 Deferred Tax - Other
313 Interest Capitalized
314 Invest Allowance Beneft
315 Dilution - Intrst Saved
316 Dilution - Divids Saved
317 Dilution - Eqty Inc Cnv
318 Accum Assesed Tax Loss
319 Accum Computd Tax Loss
320 Prior Yr Tax Adj
321 Non Cash Dividends
322 Intang Ass Written Off
323 Goodwill Written Off
324 Impairment of Investments
325 Impairment of Loans
326 Cap Pft/Loss on Fin Assts
327 Impairment of Fixed Assts
328 Cap Pft/Loss on Fix Assts
329 Pft/Loss Forex Translate
330 Pft/Loss Forex Transact
331 Pft/Loss Dispose Subsids
332 Pft/Loss Sundry Extraords
333 STC as Published
334 Non-Cash Div (Curr Year)
335 Non-Cash Div (Prev Year)

CASH FLOW ('000s)

701 Operating Profit/loss
702 Depr & Non Cash-items

703 Cash Ex Operations
704 Plus: Investment Income
705 Other Income
706 Decr/incr Work Cap

707 Decr/incr In Stock
708 Decr/incr Acc Receivable
709 Incr/decr Acc Payable
710 Incr/decr Int-free Loans

711 Csh Ex Operating Activity
712 Less: Net Int Paid/rec
713 Taxation Paid

714 Cash Available
715 Less: Ord Dividend
716 Pref Dividend

717 Net Retained Cash
718 Less: Cash Invested

719 Fixed Assets Acquired
720 Incr In Investments
721 Net Invst In Subs
722 Other Expenses/losses

723 Plus: Cash Ex Invest Acti

724 Proceeds Disp Fix Asset
725 Proceeds Disp Investmen
726 Other Proceeds

727 Cash Generated

728 Incr/decr Long Term Liab
729 Incr/decr Shrt Term Liab
730 Change In Share Capital
731 Other

732 Cash Utilised

VALUE ADDED ('000s)

760 Turnover
761 Extraordinary Items
762 Other
763 Less : Bought Mat/Serv

764 Value Added

765 Salaries & Wages
766 Interest (Net)
767 Dividends : Ordinary
768 Prefs
769 Minority
770 Taxation
771 Depreciation
772 Retention
773 Minority Interest
774 Other

775 Disburse of Value Added

776 Leasing : Property
777 Other
778 Dividends Received
779 Interest Received
780 Deferred Taxation
781 Number of Employees

SUNDRY DATA INFO

101 #Ord Shr Iss YrEnd Spl Adj
102 Nr Ord Shares Issued
103 Par Or No Par Value
110 Debtors As Surety
111 Dir Val Unlisted Invest
112 Market Val Listed Invest
113 Dir Val Uncon Subsids
114 Arrear Cum Dividends
115 Months Covered By Fin St
116 Month Of Fin Year End
117 Audit Report Qualified
118 Infl Adj Other Fix Asset
119 Infl Adj Depr Fix Asset
120 No Of Subsidiaries
121 No Of Foreign Subs
122 No Of Quoted Subs
123 Controlled By Another Co
124 Prov For Incr Repl Value
125 Pref Share Issued At Par
126 Directors Sharehold Dir
127 Directors Sharehold Ind
128 Deferred Tax Total
129 Deferred Tax For Year
130 Items Not Repr Cashflow
131 No Persons Employed
132 Stock - Raw Material
133 - Finished Goods
134 - Merchandise
135 - Consum Store
136 - Work Progress
137 - Uncompl Contrt
138 Prop Profit Ass Co's
139 Tot Res Accrued Ass Co's
140 Capital Commitments
141 Acc Deprec Land / Build
142 Lt Group Loans Advanced
143 St Group Loans Advanced
144 Head Earnings/Share
145 Lt Group Loans Received
146 St Group Loans Received
147 Notes To Statements
148 Number Of Analysts
149 Average Price Per Share
150 Jse Price Co Fin Yr End
151 Stock Valuation Method
152 Mining Assets
153 Explor Amor Exp Writ Off
154 Undeveloped Property
155 Dev Prop Less Dev Exp
156 Debtors For Prop Sold
157 Prov For Future Dev
158 Curr. Adjust. R1000 To...
162 Creditors
163 Loan Portion Of Tax
164 Bal Sheet Lifo Stock Adj
165 Inc State Lifo Stock Adj
166 Leasehold Commitments
167 Contingent Liabilities
168 Extraord Item In Tax
169 Extraord Item In Min Int
170 No Of Shares Traded
171 No Of Transactions
172 Value Of Transactions
173 Split Factor (3 Dec)
174 Month Of Stock Split

APPENDIX D – FINAL SAMPLE SELECTION

"F" = Failed
"NF" = Non-failed

The columns in the following table contain the following information:

"Ref Code"	This is the code by which companies were identified for the purposes of this study. The letter corresponds to the failed/non-failed state of the company. The arbitrarily assigned number is the same for each failed/non-failed pair (selected per a paired sample selection procedure).
"Year of F"	This is the financial year in which the company failed.
"Yrs Data"	This is the number of years of data that were available for each individual company.
"(1) Ords"	This refers to the failure criterion relatin g to the non-payment of ordinary dividends (see Chapter Thirteen).
"(2) NP"	This refers to the failure criterion relating to poor profitability as measured by net profits (see Chapter Thirteen).
"(3) NBV"	This refers to the failure criterion relating to the negative net book value of the company (see Chapter Thirteen)
"(4) NCA"	This refers to the failure criterion relating to the negative net current assets of the company (see Chapter Thirteen)
"(5) MVE"	This refers to the failure criterion relating to the fall in the market value of the equity of the company (see Chapter Thirteen)

FAILED COMPANIES

Ref Code	Short Name	Code	Company Name	Sector	Year 1st F	Yrs Data	(1) Code	(2) NP	(3) NBV	(4) NCA	(5) MVE
F1	ABACUS	ABC	Abacus Technology Holdings Ltd	Media and Entertainment	2001	3	F	F	NF	F	F
F2	AMLAC	ALC	Amlac Ltd	General Retailers	2000	3	F	F	NF	NF	NF
F3	BATECOR	BTR	Bateman Industrial Corp Ltd	Engineering and Machinery	1999	3	F	F	NF	NF	F
F4	BILLBOARD	BLL	Billboard Holdings Ltd	Media and Entertainment	2000	2	F	NF	NF	NF	F
F5	CCH	CCH	Computer Configuration Holdings Ltd	Information Technology	2000	3	F	F	NF	NF	F
F6	CHET	CET	Chet Industries Ltd	General Retailers	2000	3	F	NF	NF	NF	NF
F7	CIH	CIH	Central Information Holdings Ltd	Telecommunication Services (VCM)	2000	2	F	F	NF	F	F
F8	CORE	COR	Core Holdings Ltd	Information Technology (VCM)	2001	3	F	NF	F	NF	F
F9	CORNICK	CRK	Cornick Group Ltd	Household Goods and Textiles	1998	3	F	NF	NF	NF	F
F10	CRUX	CRX	Crux Technologies Ltd	Information Technology	2001	3	F	NF	NF	NF	F
F11	DAEWOO	DAE	Daewoo Electronics SA Ltd	Household Goods and Textiles	2001	3	F	F	F	F	F
F12	EDATA	EDT	E-Data Holdings Ltd	Information Technology (VCM)	2000	2	F	NF	NF	NF	F
F13	ESSENT	ESS	Essential Beverage Holdings Ltd	Beverages (VCM)	2000	2	F	F	NF	F	F
F14	FASHAF	FSH	Fashion Africa	General Retailers	2001	2	F	F	F	NF	F
F15	FORZA	FOZ	Forza Group Ltd	General Retailers	2001	3	F	NF	NF	NF	NF
F16	FRANSAF	FRS	Fransaf Ltd	Chemicals	1995	3	NF	F	NF	NF	F
F17	GLOHOLD	GLP	Glohold Ltd	General Retailers	2000	3	F	F	NF	NF	F
F18	GONDIE	GDL	Gondie Ltd	Media and Entertainment	2000	3	F	NF	NF	NF	NF
F19	HARWILL	HRL	Harwill Investments Ltd	Media and Entertainment	1999	3	F	F	NF	F	NF
F20	HEAVEN	HVN	Sweets From Heaven Holdings Ltd	General Retailers	2002	3	F	F	F	F	NF
F21	HICOR	HCR	Hicor Ltd	General Retailers	2001	3	F	F	NF	NF	F
F22	HOMECHOIC	HCH	Homechoice Holdings Ltd	General Retailers	2001	3	F	NF	NF	NF	F
F23	IFUSION	IFS	I-Fusion Holdings Ltd	Information Technology	2002	3	F	NF	NF	NF	NF
F24	IMPERILOG	IPG	Imperilog Ltd	General Retailers	2001	2	F	NF	NF	F	NF
F25	ITECH	ITH	Integrated Technology Holdings Ltd	Information Technology	1999	3	F	F	NF	NF	F
F26	KAROS	KAR	Karos Hotels Ltd	Leisure and Hotels	1998	3	F	NF	NF	F	F
F27	LASER	LSR	The Laser Group	Support Services	2002	3	F	F	NF	F	NF
F28	LENCO	LNC	Lenco Holdings Ltd	Diversified Industrials	2000	3	F	F	NF	F	NF
F29	MASHOLD	MKL	Mas Holdings Ltd	General Retailers	1998	3	F	F	F	NF	NF
F30	METTECH	MBT	MB Technologies Ltd	Information Technology	2001	3	F	NF	NF	NF	NF
F31	MEIJE & Z	MZG	Meije & Ziegler Ltd	Diversified Industrials	2002	3	NF	F	NF	NF	NF
F32	METKOR	MTK	Metkor Group Ltd	Diversified Industrials	1999	3	NF	NF	NF	NF	F
F33	MICOR	MIR	Micor Ltd	Transport	2000	3	NF	NF	NF	NF	F
F34	MICROLOGX	MRX	Micrologix Ltd	Information Technology (VCM)	2001	3	F	F	NF	NF	F
F35	MHH	MHH	MHH Holdings Ltd	Media and Entertainment	2002	3	F	NF	NF	NF	F
F36	MMWTECH	MMW	MMW Technology Holdings Ltd	Information Technology (VCM)	1998	2	F	NF	NF	NF	F
F37	MOLOPE	MOL	Molope Group Ltd	Support Services	1999	2	F	NF	NF	NF	NF
F38	MONEX	MNX	Monex Ltd	Leisure and Hotels	2000	3	F	NF	NF	NF	F
F39	MTRDPLS	MTR	Mthimbela Transactive Holdings Ltd	Media and Entertainment	2001	3	F	F	F	F	F
F40	NSI	NSI	National Sporting Index Ltd	Leisure and Hotels (VCM)	2001	3	F	F	NF	NF	NF
F41	OMEGA	OMA	Omega Holdings Ltd	Household Goods and Textiles	2000	3	F	F	NF	NF	F
F42	OXBRIDGE	OXB	Oxbridge Online Ltd	Information Technology	2000	3	F	F	NF	NF	F
F43	PACIFIC	PCF	Pacific Asia Investments International Ltd	General Retailers	2000	3	F	F	NF	F	F
F44	PARADIGM	PDM	Paradigm Capital Holdings Ltd	Telecommunication Services	2000	3	F	NF	NF	NF	F
F45	PERSBEL	PRB	Persbel Beleggings Bpk	Media and Entertainment	1998	2	NF	F	NF	NF	F
F46	PORTHOLD	POR	Porthold Holdings Ltd	Construction and Building Materials	2000	3	NF	NF	NF	F	F
F47	PRACTECH	PCH	Practec Technologies Ltd	Telecommunication Services (VCM)	2001	2	F	NF	F	NF	F
F48	PRIMATOY	PRT	Prima Toy & Leisure Group Ltd	General Retailers	2000	2	F	NF	NF	NF	NF
F49	PROFURN	PON	Profurn Ltd	General Retailers	2001	3	NF	NF	NF	NF	F
F50	PTH	PTH	Plant Technology Holdings Ltd	Information Technology	2001	3	F	F	NF	F	F
F51	PUTRA	PTA	Putra-Sterling Ltd	Household Goods and Textiles	1999	3	F	F	NF	NF	NF
F52	QUICKGO	QCK	Quick Holdings Ltd	Transport (DCM)	2000	3	F	F	NF	F	NF
F53	RADIOSPR	RDS	Radiospoor Technology Holdings Ltd	Telecommunication Services	1999	2	NF	NF	NF	NF	F
F54	RECTRON	RCT	Rectron Holdings Ltd	Information Technology	2001	3	F	NF	NF	NF	NF
F55	REDCORP	RTC	Redgwood Holdings Ltd	General Retailers	1998	3	F	NF	NF	NF	NF
F56	ROADCOR	ROP	Roadcorp Ltd	Transport	1998	3	F	NF	NF	F	NF
F57	SASHOLD	SSH	Stocks & Stocks Holdings Ltd	Construction and Building Materials	1997	3	NF	F	F	F	F
F58	SCHARRIG	SCG	Scharrighuisen Holdings Ltd	Construction and Building Materials	2000	3	F	NF	NF	NF	NF
F59	SEARTEC	SET	Seartec Ltd	Electronic and Electrical Equipment	2000	3	F	NF	NF	NF	F
F60	SECDATA	SDA	Securedata Solutions Ltd	Information Technology (VCM)	2000	2	F	NF	NF	NF	NF
F61	SERVEST	SRV	Servest Holdings Ltd	Support Services	2002	3	F	NF	NF	NF	F
F62	SMARTSOFT	SMT	Smartsoft Group Ltd	Information Technology (DCM)	2000	2	F	NF	NF	NF	F
F63	SOLCHEM	SLC	Solchem Investment Holdings Ltd	Media and Entertainment	2002	3	F	NF	NF	NF	F
F64	SONNDOR	SNR	Sonndor Industries Ltd	Chemicals	2000	3	F	NF	NF	NF	NF
F65	SPORT	SPT	Spine Holdings Ltd	Information Technology	2000	2	F	F	NF	NF	F
F66	SPICER	SPR	Wyespot Holdings Ltd	General Retailers	1997	3	F	F	NF	NF	F
F67	STANTRN	STT	Stantronic Group Holdings Ltd	Electronic and Electrical Equipment	1998	3	F	F	NF	NF	NF
F68	STRAND	STR	Strand Group Holdings Ltd	Chemicals	1997	3	F	F	NF	F	F
F69	STREAMWRK	SMK	Streamworks Group Ltd	Information Technology (DCM)	2002	2	F	F	NF	NF	F
F70	TELJOY	TLJ	Teljoy Holdings Ltd	Telecommunication Services	1999	3	NF	F	NF	NF	NF
F71	TILEAFRIKA	TLF	Tile Africa Holdings Ltd	General Retailers	2001	3	F	NF	NF	NF	NF
F72	TOCO	TCO	Toco Holdings Ltd	Construction and Building Materials	2000	3	F	F	NF	F	F
F73	TOLARAN	TRN	Tolaram 2000 Ltd	Household Goods and Textiles	1999	3	F	F	F	F	F
F74	TOPINTECH	TOT	Top Info Technology Holdings Ltd	Information Technology (VCM)	2000	3	F	F	NF	NF	NF
F75	TRIDELTA	TDL	Tridelta Magnet Technology Holdings Ltd	Electronic and Electrical Equipment	1999	2	F	NF	NF	NF	NF
F76	UNIFOLD	UHB	Unifold Ltd	Electronic and Electrical Equipment	2001	3	F	F	NF	NF	NF
F77	UNISPIN	UNS	Unispin Holdings Ltd	Household Goods and Textiles	1999	3	F	F	NF	NF	NF
F78	VALUECOM	VLC	Valuecom Holdings Ltd	Telecommunication Services (VCM)	2002	2	F	NF	NF	NF	F
F79	VENTEL	VTL	Venter Leisure and Commercial Trailers Ltd	Transport	2001	3	F	F	NF	NF	F
F80	WACO	WAC	Waco International Ltd	Support Services	1999	3	F	NF	NF	NF	NF
F81	WHETSTN	WTS	Whetstone Industrial Holdings Ltd	Diversified Industrials (VCM)	2000	2	F	NF	NF	F	F
F82	WINBEL	WNB	Winbel Ltd	General Retailers	2000	3	F	NF	NF	NF	NF

NON-FAILED COMPANIES

Ref Code	Short Name	Code	Company Name	Sector	(1) Code	(2) NP	(3) NBV	(4) NCA	(5) MVE
N1	CAXTON	CAT	Caxton Ctp Publishers and Printers Ltd	Media and Entertainment	NF	NF	NF	NF	NF
N2	TRUWTHS	TRU	Truworths International Ltd	General Retailers	NF	NF	NF	NF	NF
N3	INVICTA	IVT	Invicta Holdings Ltd	Engineering and Machinery	NF	NF	NF	NF	F
N4	AME	AME	African Media Entertainment Ltd	Media and Entertainment	NF	F	NF	NF	F
N5	MGX	MGX	Mgx Holdings Ltd	Information Technology	F	F	NF	NF	NF
N6	MRPRICE	MPC	Mr Price Group Ltd	General Retailers	NF	NF	NF	NF	NF
N7	HOWDEN	HWN	Howden Africa Holdings Ltd	Engineering and Machinery	NF	NF	NF	NF	NF
N8	INFOWAVE	IFW	Infowave Holdings Ltd	Information Technology (DCM)	F	NF	NF	NF	F
N9	BEARDEL	BER	Beardel Investment Corporation Ltd	Household Goods and Textiles	F	NF	NF	NF	NF
N10	EOH	EOH	Enterprise Outsourcing Holdings Ltd	Information Technology	F	NF	NF	NF	F
N11	NUWORLD	NWL	Nu-World Holdings Ltd	Household Goods and Textiles	NF	NF	NF	NF	NF
N12	LABAT	LAB	Labat Africa Ltd	Information Technology (VCM)	F	NF	NF	NF	NF
N13	AWETHU	AWT	Awethu Breweries Ltd	Beverages	F	NF	NF	NF	F
N14	NUCLICKS	NCL	New Clicks Holdings Ltd	General Retailers	NF	NF	NF	NF	NF
N15	REX-TRUE	RTO	Rex Trueform Clothing Company Ltd	General Retailers	NF	NF	NF	NF	NF
N16	SPANJAARD	SPA	Spanjaard Ltd	Chemicals	NF	NF	NF	F	NF
N17	INMINS	INM	Inmins Ltd	General Retailers	NF	NF	NF	NF	NF
N18	KGMEDIA	KGM	Kagiso Media Ltd	Media and Entertainment	F	NF	NF	NF	F
N19	SABANI	BSA	Basani Ltd	Media and Entertainment	F	NF	NF	F	NF
N20	ELLERINE	ELH	Ellerine Holdings Ltd	General Retailers	NF	NF	NF	NF	NF
N21	GLOVIL	GLI	Global Village Holdings Ltd	General Retailers	F	NF	NF	F	F
N22	BEARMAN	BRM	Bearing Man Ltd	General Retailers	NF	NF	NF	NF	NF
N23	UCS	UCS	Ucs Group Ltd	Information Technology	F	NF	NF	NF	NF
N24	CONNECT	CCT	Connection Group Holdings Ltd	General Retailers	F	NF	NF	F	NF
N25	COMPAREX	CPX	Comparex Holdings Ltd	Information Technology	NF	NF	NF	NF	NF
N26	CITYLDG	CLH	City Lodge Hotels Ltd	Leisure and Hotels	NF	NF	NF	NF	NF
N27	TRNPACO	TPC	Transpaco Ltd	Support Services	F	NF	NF	NF	NF
N28	ARGENT	ART	Argent Industrial Ltd	Diversified Industrials	NF	NF	NF	NF	NF
N29	NICTUS	NCS	Nictus Ltd	General Retailers	NF	NF	NF	NF	NF
N30	AST GROUP	AAA	Ast Group Ltd	Information Technology	F	F	NF	NF	NF
N31	BEKUNJALO	BKJ	Bekunjalo Investments Ltd	Diversified Industrials	F	F	NF	NF	F
N32	IMPERIAL	IPL	Imperial Holdings Ltd	Diversified Industrials	F	NF	NF	NF	F
N33	MOBILE	MOB	Mobile Industries Ltd	Transport	NF	NF	NF	F	F
N34	MAXTEC	MEC	Maxtec Ltd	Information Technology	F	F	NF	NF	F
N35	PRIMEM	PMA	Primedia Ltd	Media and Entertainment	F	F	NF	F	NF
N36	MUSTEK	MST	Mustek Ltd	Information Technology	F	NF	NF	NF	NF
N37	ADCORP	ACR	Adcorp Holdings Ltd	Support Services	NF	NF	NF	NF	NF
N38	GOLDREEF	GDF	Gold Reef Casino Resorts Ltd	Leisure and Hotels	F	NF	NF	NF	NF
N39	HERCOL	HCL	Heritage Collection Holdings Ltd	Media and Entertainment	F	NF	NF	NF	F
N40	STEERS	STE	Steers Holdings Ltd	Leisure and Hotels	NF	NF	NF	NF	NF
N41	BUSBY	BSB	The House Of Busby Ltd	Household Goods and Textiles	NF	NF	NF	NF	F
N42	FRONTRNGE	FRO	Frontrange Ltd	Information Technology	F	NF	NF	NF	NF
N43	BRANDCO	BRC	Brandcorp Holdings Ltd	General Retailers	NF	NF	NF	NF	F
N44	SPESCOM	SPS	Spescom Ltd	Information Technology	F	F	NF	NF	NF
N45	MNET-SS	MNS	Electronic Media Network & Supersport Intl Hld Ltd	Media and Entertainment	NF	NF	NF	NF	NF
N46	PPC	PPC	Pretoria Portland Cement Company Ltd	Construction and Building Materials	NF	NF	NF	NF	NF
N47	ZAPTRONIX	ZPT	Zaptronix Ltd	Information Technology (VCM)	F	NF	NF	NF	F
N48	WETHLYS	WET	Wetherlys Investment Holdings Ltd	General Retailers	NF	NF	NF	NF	NF
N49	FOSCHINI	FOS	Foschini Ltd	General Retailers	NF	NF	NF	NF	NF
N50	COMPCLEAR	CCL	Compu-Clearing Outsourcing Ltd	Information Technology	NF	NF	NF	NF	F
N51	PALS	PAL	Pals Holdings Ltd	Household Goods and Textiles	NF	NF	NF	NF	NF
N52	CARGO	CRG	Cargo Carriers Ltd	Transport	NF	NF	NF	F	NF
N53	JASCO	JSC	Jasco Electronics Holdings Ltd	Electronic and Electrical Equipment	NF	NF	NF	NF	NF
N54	PINNACLE	PNC	Pinnacle Technology Holdings Ltd	Information Technology	F	NF	NF	NF	F
N55	LA GROUP	LAR	La Group Ltd	General Retailers	NF	NF	NF	NF	NF
N56	PUTCO	PTC	Putco Ltd	Transport	NF	NF	NF	F	NF
N57	GROUP-5	GRF	Group Five Ltd	Construction and Building Materials	NF	NF	NF	NF	F
N58	CONCOR	CNC	Concor Ltd	Construction and Building Materials	NF	NF	NF	NF	F
N59	CONTROL	CNL	Control Instruments Group Ltd	Electronic and Electrical Equipment	NF	NF	NF	NF	NF
N60	DCENTRIX	DCT	Datacentrix Holdings Ltd	Information Technology	F	F	NF	NF	NF
N61	ENSERVE	ENV	Enviroserv Holdings Ltd	Support Services	NF	NF	NF	NF	NF
N62	SOFTLINE	SFT	Softline Ltd	Information Technology	F	F	NF	NF	NF
N63	MONEYWB	MNY	Money Web Holdings Ltd	Media and Entertainment	F	F	NF	NF	F
N64	CHEMSERVE	CHE	Chemical Services Ltd	Chemicals	NF	NF	NF	NF	NF
N65	DATATEC	DTC	Datatec Ltd	Information Technology	F	NF	NF	NF	NF
N66	WINHOLD	WNH	Winhold Ltd	General Retailers	NF	NF	NF	NF	NF
N67	PASDEC	PSC	Pasdec Resources SA Ltd	Electronic and Electrical Equipment	F	F	NF	NF	NF
N68	OMNIA	OMN	Omnia Holdings Ltd	Chemicals	NF	NF	NF	NF	NF
N69	ERP.COM	ERP	Erp.Com Holdings Ltd	Information Technology	F	NF	NF	NF	NF
N70	MTN	MTN	Mtn Group Ltd	Telecommunication Services	NF	NF	NF	NF	NF
N71	MATHOMO	MTO	Mathomo Group Ltd	General Retailers	F	F	F	NF	F
N72	MASONITE	MAS	Masonite (Africa) Ltd	Construction and Building Materials	NF	NF	NF	NF	NF
N73	ADONIS	ADD	Adonis Knitwear Holdings Ltd	Household Goods and Textiles	NF	NF	NF	NF	NF
N74	APPLIED	APL	Net 1 Applied Technology Holdings Ltd	Information Technology	F	NF	NF	NF	NF
N75	GRINTEK	GNK	Grintek Ltd	Electronic and Electrical Equipment	NF	NF	NF	NF	NF
N76	DELTA	DEL	Delta Electrical Industries Ltd	Electronic and Electrical Equipment	NF	NF	NF	NF	NF
N77	AMAPS	AMA	Amalgamated Appliance Holdings Ltd	Household Goods and Textiles	NF	NF	NF	NF	NF
N78	STELLA	SLL	Stella Vista Technologies Ltd	Information Technology (DCM)	F	NF	NF	NF	F
N79	VALUE	VLE	Value Group Ltd	Transport	F	NF	NF	NF	NF
N80	REBSERV	RBV	Rebserve Holdings Ltd	Support Services	NF	NF	NF	NF	NF
N81	MT-EAGLE	MTE	Mvelinga Societe Anonyme	Diversified Industrials	NF	NF	NF	F	NF
N82	AF-&-OVER	AOO	African & Overseas Enterprises Ltd	General Retailers	NF	NF	NF	NF	NF

APPENDIX E – RATIO CALCULATIONS

This appendix consists of the following two parts:

- **Appendix E.1**: Method of calculation for all ratios.
- **Appendix E.2**: Abbreviations for inputs into calculations in E.1.

Appendix E.1. Method of calculation for all ratios.

REF #	DESCRIPTION	CALCULATION	YRS RATIO UTILISED
LIQUIDITY RATIOS			
L1	Current Ratio	CA / CL	1,2,3
L2	Quick Ratio	(CA - Inv) / CL	1,2,3
L3	Cash Ratio	(Cash + Rec) / CL	1,2,3
L4	Non-Group Current Ratio	(CA - SGR) / (CL - SGP)	1,2,3
L5	Non-Group Quick Ratio	(QA - SGR) / (CL - SGP)	1,2,3
L6	Cash to Total Assets Ratio	Cash / TA	1,2,3
L7	Receivables Turnover	((S / Mths) x 12) / Rec	1,2,3
L8	Inventory Turnover	((S / Mths) x 12) / Inv	1,2,3
L9	Payables Turnover	((S / Mths) x 12) / Pay	1,2,3
L10	Cash Conversion Cycle	(365 / L7) + (365 / L8) + (365 / L9)	1,2,3
L11	Declaration of Preference Share Dividend	+1 = Preference shares issued and dividend declared 0 = No preference shares in issue -1 = Preference shares issued and dividend not declared	1,2,3
OPERATING EFFICIENCY RATIOS			
E1	Total Asset Turnover	((S / Mths) x 12) / TA	1,2,3
E2	Fixed Asset Turnover	((S / Mths) x 12) / FA	1,2,3
E3	Equity Turnover	((S / Mths) x 12) / BVE	1,2,3

OPERATING PROFITABILITY RATIOS			
P1	Operating Profit Margin	EBIT / S	1,2,3
P2	Net Profit Margin	NI / S	1,2,3
P3	Return on Total Capital	(((NI + IE) / Mths) x 12) / TC	1,2,3
P4	Return on Total Equity	((NI / Mths) x 12) / BVE	1,2,3
P5	Return on Ordinary Equity	(((NI - PD) / Mths) x 12) / OSI	1,2,3
P6	Basic EPS	(EPS / Mths) x 12	1,2,3
P7	Headline EPS	(HEPS / Mths) x 12	1,2,3
SOLVENCY RATIOS			
S1	Debt Equity Ratio	LTD / BVE	1,2,3
S2	Long Term Debt to Assets Ratio	LTD / TA	1,2,3
S3	Interest Bearing Debt to Asset Ratio	IBD / TA	1,2,3
S4	External Long Term Debt to Asset Ratio	(LTD - LGP) / TA	1,2,3
S5	Financial Leverage	TA / BVE	1,2,3
S6	Total Debt to Assets Ratio	TD / TA	1,2,3
S7	Total Debt and Contingencies to Assets Ratio	(TD + Cont) / TA	1,2,3
S8	Total Commitments to Assets Ratio	(TD + Cont + CC + LC) / TA	1,2,3
S9	External Debt to Assets Ratio	(TD - SGP - LGP) / (TA - SGR - LGR)	1,2,3

CASH FLOW RATIOS			
C1	Interest Coverage Ratio (Inverse)	IE / (EBIT + II)	1,2,3
C2	Fixed Charge Coverage Ratio (Inverse)	(IE + LP + (PD / 1 - TaxRate)) / (EBIT + LP + II)	1,2,3
C3	Interest Expense to Cash Flow Ratio	IE / CF	1,2,3
C4	Cash Flow to Long Term Debt Ratio	((CF / Mths) x 12) / LTD	1,2,3
C5	Cash Flow to Interest Bearing Debt (Inverse)	((CF / Mths) x 12) / IBD	1,2,3
C6	Cash Flow to Total Debt Ratio	((CF / Mths) x 12) / TD	1,2,3
C7	Cash Flow from Operations to Total Debt	((CFO / Mths) x 12) / TD	1,2,3
C8	Change in Long Term Debt to Long Term Debt	((CFLTD / Mths) x 12) / LTD	1,2,3
C9	Change in Total Debt to Total Debt	(((CFLTD + CFSTD + CFAP) / Mths) x 12) / TD	1,2,3
C10	Proceeds on Share Issue to Total Assets	CFSP / TA	1,2,3
C11	Cash Invested in Investing Activities to Fixed Assets	((CFIA / Mths) x 12) / FA	1,2,3
C12	Proportion of Dividend that is Non-Cash	NCDiv / Div	1,2,3
MARKET RATIOS			
M1	Trading Turnover	((NST / Mths) x 12) / Ords	1,2,3
M2	Total Dividend Yield	((DPS / Mths) x 12) / SP	1,2,3
M3	Earnings Yield	((EPS / Mths) x 12) / SP	1,2,3
M4	Market Value of Equity to Total Debt	MVE / TD	1,2,3
M5	Price Book Ratio	MVE / OSI	1,2,3
M6	Capital to EBITDA Ratio	(MVE + Debt) / (EBITDA / Mths x 12)	1,2,3
M7	Excess Return (1 Fin Yr)	(SP2 / SF2 - SP1 / SF1) / (SP2 / SF2) - JSE ALSI Return f	1,2
M8	Excess Return (2 Fin Yrs)	(SP3 / SF3 - SP1 / SF1) / (SP3 / SF3) - JSE AlSI Return f	1

RISK ANALYSIS RATIOS			
R1	Business Risk (3yrs)	(Stdev EBIT) / (Avg EBIT)	1
R2	Sales Variability (3yrs)	(Stdev S) / (Avg S)	1
R3	Operating Leverage (3yrs)	((EBIT3 - EBIT1) / EBIT1) / ((S3 - S1) / S1)	1
R4	Business Risk (2yrs)	(Stdev EBIT) / (Avg EBIT)	2
R5	Sales Variability (2yrs)	(Stdev S) / (Avg S)	2
R6	Operating Leverage (2yrs)	((EBIT3 - EBIT1) / EBIT1) / ((S3 - S1) / S1)	2
SIZE AND GROWTH RATIOS			
G1	Log(Assets)	Log (TA)	1,2,3
G2	Log(Market Capitalisation)	Log (MVE)	1,2,3
G3	Retained Income to Total Assets Ratio	RI / TA	1,2,3
G4	Retention Ratio	1 - (Div / EAT)	1,2,3
G5	Growth Rate	G4 / ROE	1,2,3
G6	Capital Commitments to Fixed Assets Ratio	CC / FA	1,2,3
NON-FINANCIAL RATIOS			
N1	Directors Shareholding	DSH / Ords	1,2,3
N2	Auditors Report Qualification	0 = Audit report qualified 1 = Audit report not qualified	1,2,3
N3	Controlled by Another Company	0 = Not controlled by another company 1 = Controlled by another company	1,2,3
N4	Number of Years Listed	(Year of failure OR 2003) - (Year of listing)	1
N5	Number of Years in Existence	(Year of failure OR 2003) - (Year of commencement)	1

Appendix E.2. Abbreviations for inputs into calculations in E.1.

Symbol	Description	#
A	Total Assets (excluding Intangibles)	50
AR	Audit Report Qualification	117
BE	Bottomline Earnings	80
BVE	Book Value of Total Shareholders Equity (OSE+CPref)	-
CA	Current Assets	34
Cash	Cash and Near Cash	37
CC	Capital Commitments	140
CF	Cash Generated from Operations	711
CFAP	Cash Flow from Accounts Payable	709
CFIA	Cash Invested in Investing Activities	718
CFLTD	Cash Flow from Long Term Debt	728
CFO	Cash Flows from Operating Activities	717
CFSP	Proceeds on Share Issue	730
CFSTD	Cash Flow from Short Term Debt	729
CL	Current Liabilities	41
Cont	Contingent Liabilities	167
CPref	Convertible Preference Share	11
DDSH	Direct Directors' Shareholding	126
Dep	Depreciation	88
Div	Ordinary Dividend Declared (Cash and Non-Cash)	81
DPS	Dividends Per Share	87
DSH	Director's Shareholding (DDSH + IDSH)	-
EAT	Earnings after Tax	71
EBIT	Earnings before Interest, Tax and Intangibles Written Off (OP+IWO)	-
EPS	Bottomline Earnings Per Share	305
FA	Fixed Assets	23
FE	Income Statement Finance Charge	66
GWO	Goodwill Written Off	323
HC	Control by Holding Company	123
HEPS	Headline Earnings Per Share	306
IA	Intangible Assets	25
IBD	Interest Bearing Debt (IBLD+IBSD)	-
IBLD	Interest Bearing Long Term Debt	20
IBSD	Interest Bearing Short Term Debt	45
IC	Interest Capitalised	313
IDSH	Indirect Directors' Shareholding	127
IE	Interest Expense (FE+IC)	-
II	Interest Income	64
Intan	Intangible Assets excluding Patents and Trademarks (IA-PT)	-
Inv	Inventory	35
IPref	Irredeemable Preference Shares	9
IWO	Intangibles Written Off (GWO+OIWO)	-
LC	Leasehold Commitments	166
LGP	Long Term Group Payables	145

Symbol	Description	#
LGR	Long Term Group Receivables	142
LP	Lease Payments (LPB+LPO)	-
LPB	Lease Payments - Buildings	301
LPO	Lease Payments - Other	302
LTD	Long Term Debt (LTL+RPref+IPref)	-
LTL	Long Term Liabilities	16
Mths	Months covered by FS	115
MVE	Market Capitalisation (SP*Ords/100)	-
NCDiv	Non-Cash Dividend	321
NI	Bottom Line Net Income excluding Intangibles Written Off (BE+IWO)	-
NST	Number of Shares Traded during the period	170
OIWO	Other Intangibles Written Off	322
OP	Operating Profit	63
Ords	Number of Ordinary Shares in Issue	102
OSE	Ord Sholders Equity less Intangibles (excluding P&T) (OSI-Intan)	-
OSI	Ordinary Shareholders Interest	1
Pay	Trade Creditors	42
PD	Preference Dividend Declared	72
PT	Patents and Trademarks	27
QA	Quick Assets (CA-Inv)	-
Rec	Trade Debtors	36
RI	Retained Income at the End of Period	5
RPref	Redeemable Preference Shares	10
S	Sales	60
SF	Split Factor	173
SGP	Short Term Group Payables	146
SGR	Short Term Group Receivables	143
SP	Share Price at Period End	150
SPX	Share Price at Year X End	150
TA	Total Assets less Intangibles (excluding Patents and Trademarks) (A+PT)	-
TaxRate	Effective Tax Rate	[illegible]
TC	Capital Employed less Intangibles	49
TD	Total Debt (LTD+CL)	-

APPENDIX F – SAMPLE DATA ITEM VALUES

This appendix presents the values for all the ratios calculated and used as inputs in the feature subset selection process.

The appendix consists of the following components:

- **Appendix F.1**: Ratio input vectors for the failed companies calculated using the data from the final year prior to failure.
- **Appendix F.2**: Ratio input vectors for the failed companies calculated using the data two years prior to failure.
- **Appendix F.3**: Ratio input vectors for the failed companies calculated using the data three years prior to failure
- **Appendix F.4**: Ratio input vectors for the non-failed companies paired with those ratios presented in Appendix F.1.
- **Appendix F.5**: Ratio input vectors for the non-failed companies paired with those ratios presented in Appendix F.2.
- **Appendix F.6**: Ratio input vectors for the non-failed companies paired with those ratios presented in Appendix F.3.

This appendix makes reference to each company using its code as assigned in Appendix D. The ratios have been abbreviated and can be interpreted as follows (note that these ratios have not yet been normalised, but that missing data has been filled in as described in Chapter Sixteen):

This makes reference to the variable code allocated in Appendix E.1.

L1 (1)

This refers to the year of the data. A "(1)" refers to variables relating to data from the financial year one year prior to failure.

The company input vectors have been highlighted if that company has only two years of data available prior to failure (or, in the case of non-failed companies, the company had been paired with a failure that has only two years of available data).

30 Year 1 Financial Variables Selected (Administration Rates (Year 1986 to Present))

Ref Code	Firm Name	Variables (values)
F1	ABACUS	[illegible]
F2	AMLAD	[illegible]
F3	BATECOM	[illegible]
F4	BISARBO	[illegible]
F5	DC-	[illegible]
F6	CHST	[illegible]
F7	CM	[illegible]
F8	DOME	[illegible]
F9	CORNICK	[illegible]
F10	CRUX	[illegible]
F11	DAEWOO	[illegible]
F12	ENATA	[illegible]
F13	ESSENT	[illegible]
F14	FASHAF	[illegible]
F15	FORZA	[illegible]
F16	FRANBAP	[illegible]
F17	GLOHOLD	[illegible]
F18	GUHOLE	[illegible]
F19	HARVILL	[illegible]
F20	HEAVEN	[illegible]
F21	HICORL	[illegible]
F22	HOMECHOIC	[illegible]
F23	IFUSION	[illegible]
F24	IMPEALOG	[illegible]
F25	ITECH	[illegible]
F26	KAROS	[illegible]
F27	LASER	[illegible]
F28	LEMCO	[illegible]
F29	MASHOLD	[illegible]
F30	METECH	[illegible]
F31	METJE-S-E	[illegible]
F32	METKOR	[illegible]
F33	MICOR	[illegible]
F34	MICROLOGIX	[illegible]
F35	MIDA	[illegible]
F36	MMATECH	[illegible]
F37	MOLOPE	[illegible]
F38	MRMES	[illegible]
F39	NITROPLS	[illegible]
F40	MDI	[illegible]
F41	OMEGA	[illegible]
F42	OLBRIDGE	[illegible]
F43	PACIFIC	[illegible]
F44	PARADION	[illegible]
F45	PERDISEL	[illegible]
F46	PORTHLD	[illegible]
F47	PRASTECH	[illegible]
F48	PRIMATDY	[illegible]
F49	PROFURM	[illegible]
F50	PTH	[illegible]
F51	PUTRA	[illegible]
F52	QUICKUO	[illegible]
F53	RAD-OSPR	[illegible]
F54	REDTRON	[illegible]
F55	RBTCORP	[illegible]
F56	ROADCOR	[illegible]
F57	USSHOLD	[illegible]
F58	SCHARIS	[illegible]
F59	SEAMTEC	[illegible]
F60	SECCATA	[illegible]
F61	SERVFST	[illegible]
F62	SMARSOFT	[illegible]
F63	SOICHEM	[illegible]
F64	SONDOR	[illegible]
F65	SPORT	[illegible]
F66	SPICER	[illegible]
F67	STARTRY	[illegible]
F68	STRAND	[illegible]
F69	STREAMARK	[illegible]
F70	TELJOY	[illegible]
F71	TLEXTRINA	[illegible]
F72	TOCO	[illegible]
F73	TOLARAN	[illegible]
F74	TOP-TECH	[illegible]
F75	TRIDELTA	[illegible]
F76	UNIPOLE	[illegible]
F77	UNISPIN	[illegible]
F78	VALUECOM	[illegible]
F79	VERTEL	[illegible]
F80	WADO	[illegible]
F81	WHETSTN	[illegible]
F82	WINGEL	[illegible]
	Max	[illegible]
	Average	[illegible]

At Year 1 Potential Variables Selected (Formal Variable Rank (Year Prior to Failure))

Company Name	Ref Code
ABACUS	F1
AMTEC	F2
BATEGUR	F3
GILSBANE	F4
CCH	F5
CHST	F6
CIII	F7
CORE	F8
DORAICK	F9
CRSX	F10
DAEWOO	F11
EDATA	F12
ESSENT	F13
FASHAT	[illegible]
FORZA	[illegible]
FRANJAF	[illegible]
GLOHOLE	[illegible]
GUNBLT	[illegible]
HARMILL	[illegible]
HEAVEN	[illegible]
HICORL	[illegible]
HOMECHOIC	[illegible]
HUSKIN	[illegible]
IMPERILCO	[illegible]
ITECH	F26
KAMDO	[illegible]
LASER	[illegible]
LEVDO	[illegible]
MASHOLD	[illegible]
MBTECH	[illegible]
METJE & Z	[illegible]
METROX	[illegible]
MCOR	[illegible]
MICROLOGIA	[illegible]
MIRH	[illegible]
MMATECH	[illegible]
MOLOPE	[illegible]
MONEX	[illegible]
MTROPLD	[illegible]
NBI	[illegible]
OMEGA	[illegible]
OABRIDGE	[illegible]
PACIFIC	[illegible]
PARAGIGM	[illegible]
PERRREI	[illegible]
PORTHLD	[illegible]
PRANTECH	[illegible]
PRIMATOY	[illegible]
PROFURM	[illegible]
PTH	[illegible]
PUTRA	[illegible]
QUICKCO	[illegible]
RADIOSER	[illegible]
RECTRON	[illegible]
RSTCORP	[illegible]
ROACCOR	[illegible]
SASBDIO	[illegible]
SCNARIG	[illegible]
SCARTEC	[illegible]
SECDATA	[illegible]
SERVEST	[illegible]
SMACRAFT	[illegible]
SOTCIRTM	[illegible]
SONGOT	[illegible]
SPORT	[illegible]
SPICER	[illegible]
STANTRN	[illegible]
STRAND	[illegible]
STREAMARK	[illegible]
TELJOY	[illegible]
TELEAFRIKA	[illegible]
TOGO	[illegible]
TOLARAM	[illegible]
TOP-TECH	[illegible]
TRIDELTA	[illegible]
UNIHOLD	[illegible]
UNISPIN	[illegible]
VALLECON	[illegible]
VENTEL	[illegible]
WACO	[illegible]
WHETSTN	[illegible]
WINBEL	F82

Appendix F.2.

Ref Code	Firm Name	[illegible]
F1	AMALGN	[illegible]
F2	ANSLAC	[illegible]
F3	BATECOR	[illegible]
F4	EC BOARD	[illegible]
F5	CSM	[illegible]
F6	DMST	[illegible]
F7	CP	[illegible]
F8	CORE	[illegible]
F9	CORRICK	[illegible]
F10	CRUX	[illegible]
F11	DAEWOO	[illegible]
F12	FOATA	[illegible]
F13	ESSFLT	[illegible]
F14	FAMMAP	[illegible]
F15	FORZA	[illegible]
F16	FRANBAP	[illegible]
F17	GLOHOLD	[illegible]
F18	DURALF	[illegible]
F19	HARRILL	[illegible]
F20	HEAVEN	[illegible]
F21	HOORL	[illegible]
F22	HONGHOIG	[illegible]
F23	FUSION	[illegible]
F24	HYPERLOG	[illegible]
F25	ITECH	[illegible]
F26	KAMDB	[illegible]
F27	LABER	[illegible]
F28	LEMCO	[illegible]
F29	MASHOLD	[illegible]
F30	METECH	[illegible]
F31	METJEAS	[illegible]
F32	MECROR	[illegible]
F33	MECOR	[illegible]
F34	MICROLOGIX	[illegible]
F35	MMM	[illegible]
F36	MONOTECH	[illegible]
F37	MOLDPE	[illegible]
F38	MONCK	[illegible]
F39	MTPORLS	[illegible]
F40	MST	[illegible]
F41	OMEGA	[illegible]
F43	OXBRIDGE	[illegible]
F45	PACIFIC	[illegible]
F46	PARAGON	[illegible]
F48	PENGEL	[illegible]
F54	PORTHL	[illegible]
F42	PRADTECH	[illegible]
F45	PRIMATOY	[illegible]
F49	PROFURN	[illegible]
F50	PTM	[illegible]
F51	PUTRA	[illegible]
F52	QUIGNCO	[illegible]
F53	RADIOSPR	[illegible]
F54	RECTRON	[illegible]
F55	RETCORP	[illegible]
F56	ROADCOR	[illegible]
F57	RAROLS	[illegible]
F58	SCHARD	[illegible]
F59	SEARTEC	[illegible]
F60	SEDATA	[illegible]
F61	SERVEST	[illegible]
F62	SYNCSOFT	[illegible]
F63	SOLCHEM	[illegible]
F64	SANDOR	[illegible]
F65	SPORT	[illegible]
F66	SHOER	[illegible]
F67	STARTRN	[illegible]
F68	STRANC	[illegible]
F69	STRESSMARK	[illegible]
F70	TELJOY	[illegible]
F71	TLEARNA	[illegible]
F72	TOGO	[illegible]
F73	TOLARAN	[illegible]
F74	TOPTECH	[illegible]
F75	TRIDELTA	[illegible]
F76	UNIHOLD	[illegible]
F77	UNIPAK	[illegible]
F78	VALUEPCOM	[illegible]
F79	VENTEL	[illegible]
F80	WACC	[illegible]
F81	WINETECH	[illegible]
F83	WINGEL	[illegible]
	Max	[illegible]
	Average	[illegible]

All Years Potential Variables Selected (Partial, Variable Name (Year Prior to Failure))

Variables (all columns)	Short Name	Fail Code
[illegible]	ABACUS	F1
[illegible]	AMLAG	F2
[illegible]	BATEDOR	F3
[illegible]	BILBRARD	F4
[illegible]	CCH	F5
[illegible]	C-ET	F6
[illegible]	C--	F7
[illegible]	CORC	F8
[illegible]	CORSICH	F9
[illegible]	CRUX	F10
[illegible]	DAFACO	F11
[illegible]	EDATA	F12
[illegible]	ESSERT	F13
[illegible]	FACHAP	F14
[illegible]	FORZA	F15
[illegible]	FRANCAP	F16
[illegible]	GLDHOLD	F17
[illegible]	GWYDLE	F18
[illegible]	HARWALL	F19
[illegible]	HEAVEN	F20
[illegible]	HIGCRL	F21
[illegible]	HOMECROWG	F22
[illegible]	IFUSIDN	F23
[illegible]	IMPRILON	F24
[illegible]	ITECH	F25
[illegible]	KAROB	F26
[illegible]	LAHER	F27
[illegible]	LFNDO	F28
[illegible]	MAGHOLD	F29
[illegible]	MBTECH	F30
[illegible]	METJE-A-Z	F31
[illegible]	METKOR	F32
[illegible]	MICCA	F33
[illegible]	MICROLDRX	F34
[illegible]	MIHH	F36
[illegible]	MMATECH	F36
[illegible]	MOLOPE	F37
[illegible]	MONCR	F38
[illegible]	MTHOFLD	F39
[illegible]	MUI	F40
[illegible]	OMEGA	F41
[illegible]	OXBRIDGE	F42
[illegible]	PACIFIC	F43
[illegible]	PARADIGM	F44
[illegible]	PERSNET	F45
[illegible]	PORTHLD	F46
[illegible]	PRADTECH	F47
[illegible]	PRIMATOP	F48
[illegible]	PROFURN	F49
[illegible]	PTH	F50
[illegible]	PUTRA	F51
[illegible]	QUICKCO	F52
[illegible]	RADIOSPA	F53
[illegible]	RECTRON	F54
[illegible]	RETCORP	F55
[illegible]	ROADCOR	F56
[illegible]	SABHOLD	F57
[illegible]	SCHARIG	F58
[illegible]	SEARTES	F59
[illegible]	SECDATA	F60
[illegible]	SERVEST	F61
[illegible]	SMACSOFT	F62
[illegible]	SOLCHEN	F63
[illegible]	SONDOR	F64
[illegible]	SPORT	F65
[illegible]	SPICER	F66
[illegible]	STANTRN	F67
[illegible]	STRAND	F68
[illegible]	STREAMRK	F69
[illegible]	TELJOY	F70
[illegible]	TILECRA	F71
[illegible]	TOCO	F72
[illegible]	TOLARAN	F73
[illegible]	TOP-TECH	F74
[illegible]	TRIDELTA	F75
[illegible]	UNIHOLD	F76
[illegible]	UNIOPIN	F77
[illegible]	VALUECOM	F78
[illegible]	VENTEL	F79
[illegible]	WAGO	F80
[illegible]	WRESTETH	F81
[illegible]	WABEL	F82
[illegible]	Max	
[illegible]	Average	

Appendix F.3.

Ref Code	Firm Name	Values
F1	ABACUS	[illegible]
F2	AIRLAC	[illegible]
F3	BATSCOM	[illegible]
F4	BILBOARD	[illegible]
F5	CEN	[illegible]
F6	CHET	[illegible]
F7	CIN	[illegible]
F8	CORE	[illegible]
F9	CORNICA	[illegible]
F10	DTRS	[illegible]
F11	DAEWOO	[illegible]
F12	EDATA	[illegible]
F13	EDENT	[illegible]
F14	FAIRFAX	[illegible]
F15	FORZA	[illegible]
F16	FRANCAP	[illegible]
F17	GLOWBLD	[illegible]
F18	GUNDLE	[illegible]
F19	HASWILL	[illegible]
F20	HEAVEN	[illegible]
F21	HICOAL	[illegible]
F22	HOMECHOIC	[illegible]
F23	IFUSION	[illegible]
F24	IVPENRING	[illegible]
F25	ITECH	[illegible]
F26	KAZOS	[illegible]
F27	LAFIS	[illegible]
F28	LANDB	[illegible]
F29	MAXHOLD	[illegible]
F30	MATECH	[illegible]
F31	METZE&J	[illegible]
F32	METKOR	[illegible]
F33	NICOL	[illegible]
F34	NICOLCORE	[illegible]
F35	NIAH	[illegible]
F36	[illegible]	[illegible]
F37	NCLAPS	[illegible]
F38	OTHERS	[illegible]
F39	OTESPLS	[illegible]
F40	NTI	[illegible]
F41	ORECA	[illegible]
F42	OSBSIDE	[illegible]
F43	PACIFIC	[illegible]
F44	PARADISE	[illegible]
F45	PERSONAL	[illegible]
F46	PORTPLS	[illegible]
F47	PRACTION	[illegible]
F48	PRIMATOY	[illegible]
F49	PROFUSION	[illegible]
F50	PTH	[illegible]
F51	PUTRA	[illegible]
F52	QJOKOD	[illegible]
F53	RADIOPR	[illegible]
F54	REDTWO	[illegible]
F55	REICORP	[illegible]
F56	RONDCOA	[illegible]
F57	SAGHOLD	[illegible]
F58	SCHAMG	[illegible]
F59	SMARTEC	[illegible]
F60	SNOSTA	[illegible]
F61	SORUSH	[illegible]
F62	SVAGFIR	[illegible]
F63	SOLGHES	[illegible]
F64	SONDPN	[illegible]
F65	SPORT	[illegible]
F66	SPICES	[illegible]
F67	STANTON	[illegible]
F68	STRAND	[illegible]
F69	STREAMWORK	[illegible]
F70	TELJOS	[illegible]
F71	TELEPHONE	[illegible]
F72	TODO	[illegible]
F73	TOLARAN	[illegible]
F74	TOPTECH	[illegible]
F75	TRONSIA	[illegible]
F76	UNIMOLE	[illegible]
F77	UNISPAN	[illegible]
F78	VALUECOM	[illegible]
F79	VENTEL	[illegible]
F80	WACO	[illegible]
F81	WHETSTN	[illegible]
F82	WINGEL	[illegible]
	Max	[illegible]
	Average	[illegible]

Appendix F.4.

Ref Code	Short Name	All Year 1 Potential Variables Selected [Format: Variable Name (Year Prior to Failure)]																													
		L1(1)	L2(1)	L3(1)	L4(1)	L5(1)	L6(1)	L7(1)	L8(1)	L9(1)	L10(1)	L11(1)	E1(1)	E2(1)	E3(1)	F1(1)	P2(1)	P3(1)	P4(1)	P5(1)	P6(1)	P7(1)	S1(1)	S2(1)	S3(1)	S4(1)	S5(1)	S6(1)	S7(1)	S8(1)	S9(1)
N1	CAXTON	3.577536	2.991566	2.935177	3.583159	2.99629	0.303729	4.551226	8.542422	6.523617	82.49763	-1	0.995092	3.354699	1.573958	0.108279	0.076783	0.092364	0.120653	8.120653	469	475	0.000422	0.000267	0	0.000267	1.58172	0.178236	0.178236	0.188688	0.177957
N2	TRUWTHS	3.324997	2.810776	2.729606	2.982448	2.463748	0.136697	3.516787	14.33196	15.79396	106.0664	0	1.585495	14.42818	2.413506	0.002639	-0.00514	-0.00117	-0.01241	-0.01241	-2.1	37.3	0	0	0	0	1.522242	0.215135	0.215135	1.435415	0.23163
N3	INVICTA	2.697147	1.822051	1.822051	2.697147	1.822051	0.37703	5.66509	3.560973	4.494523	65.7196	0	1.259237	28.28036	1.978905	0.093744	0.065912	0.16953	0.170012	0.170012	45	45	0.002394	0.001523	0.020218	0.001523	1.571511	0.330444	0.448134	0.463949	0.330444
N4	AME	0.668236	0.726634	0.667639	0.668236	0.726634	0.049665	5.949911	28.3124	5.231364	5.445763	0	3.015966	15.3496	-32.7012	-0.08532	-0.10639	-0.36599	3.446249	3.446249	-4.2	0.7	-2.66147	0.245462	0.230249	0.245462	-10.6427	1.054923	1.054923	1.066663	1.054923
N5	MGX	1.697196	1.469061	1.469061	1.697196	1.469061	0.264884	3.024669	10.6635	4.710455	76.78511	0	0.607512	2.751212	1.703914	0.309639	0.148446	0.196186	0.25294	0.25294	147.3	147.3	0.345079	0.163536	0.103623	0.163536	2.110076	0.52071	0.52071	0.769235	0.52071
N6	MRPRICE	1.710243	0.067652	0.569735	1.710338	9.66769	0.076609	15.50297	5.626426	9.414582	47.41963	1	2.227605	13.35473	4.601746	0.057017	0.044203	0.16169	0.212252	0.212218	43.3	43.5	0.001715	0.000796	0.000644	0.000796	2.155372	0.37471	0.432366	0.527754	0.37469
N7	HOWDEN	1.765001	1.204666	1.145439	1.774282	1.209232	0.145489	4.377423	6.419996	9.125281	100.2372	0	1.615065	9.667581	3.263778	0.022153	0.013497	0.226003	0.044052	0.044052	7.6	7.6	0	0	0	0	2.020608	0.449127	0.463528	0.471905	0.448453
N8	INFOWAVE	1.725775	1.725775	1.659655	1.725775	1.725775	0.399757	7.402571	59	5.087654	-16.2487	0	1.902452	21.28531	3.424949	0.09373	0.064618	0.204287	0.221316	0.221316	1.6	1.9	0.087964	0.046661	0	0.046661	1.800262	0.444531	0.444531	0.65519	0.444531
N9	SEARDEL	1.587725	0.601752	0.601752	1.587725	0.601752	0.007566	7.674313	4.539993	5.518673	61.61677	0	1.834036	8.373461	3.82261	0.065473	0.052765	0.239661	0.201711	0.201711	63.2	34	0.145612	0.069955	0.139594	0.069955	2.064369	0.479676	0.482599	0.495663	0.479676
N10	EOH	1.370573	1.352204	1.266714	1.370573	1.352204	0.51294	5.414242	143.7504	35.96374	69.61045	0	1.734147	39.05699	5.355623	0.150691	0.125633	0.373409	0.672644	0.672644	20.2	20.5	0.011591	0.003753	0.002722	0.003753	3.066333	0.660506	0.660506	0.817417	0.660506
N11	NUWORLD	2.669769	1.837018	1.837018	2.669769	1.837018	0.175126	5.443795	6.573567	7.627267	61.7669	0	2.379574	31.48037	3.7863	0.057564	0.036202	0.172624	0.143679	0.143679	186.4	156.4	0.021676	0.013695	0.013776	0.013695	1.562763	0.346977	0.472429	0.525077	0.346977
N12	LABAT	2.066579	1.924036	1.893685	2.066579	1.924036	0.00021	0.821762	10.76509	16.84036	458.6893	0	0.733583	29.59946	2.216129	0.05784	-0.03767	0.093127	-0.06355	-0.06355	-33.9273	-33.9273	0.020266	0.006709	0.439217	0.006709	3.023775	0.476152	0.476152	0.724692	0.476152
N13	AWETHU	0.95831	0.73476	0.56916	0.95831	0.73476	0.056666	10.30166	20.34543	5.416617	-13.9991	0	2.018146	4.046753	925.9536	-0.26693	-0.32067	-0.97537	-296.637	-296.637	-21.2	-22.1	254.2093	0.554057	0.569944	0.554057	458.814	0.99782	0.99782	0.99782	0.99782
N14	NUCLICKS	1.415549	0.174406	0.174406	1.415549	0.174406	0.009691	49.6936	4.66456	6.664666	26.23552	0	2.293294	6.020614	4.653417	0.061693	0.040632	0.175942	0.189078	0.189078	52.2	60.3	0.258402	0.127346	0.145363	0.127346	2.029141	0.507161	0.511646	1.165603	0.507161
N15	REX-TRUE	5.412644	2.036419	2.036362	5.515051	2.076909	0.168406	11.28109	2.969202	12.96976	127.1968	1	1.471562	8.37503	1.727669	-5.3E-05	0.009063	0.015966	0.015642	0.015643	12.3	12.3	0.001721	0.001466	0	0.001466	1.174174	0.148336	0.148336	0.157911	0.148616
N16	SPANJAARD	8.958684	0.891594	0.602693	0.958684	9.691594	0.005032	5.012986	11.15505	5.468529	38.7862	0	1.899965	6.991704	6.812995	0.062265	0.013615	0.24635	0.092756	0.092756	7.9	10.8	0.270202	0.075234	0.336952	0.075234	3.591466	0.711456	0.711456	0.711456	0.711456
N17	INMNS	1.57154	0.983993	0.983993	1.57154	0.983993	0.020565	6.018992	8.70338	8.303608	40.35563	0	3.186532	27.78121	7.450982	0.047119	0.037716	0.299394	0.281034	0.281034	27.8	19.2	0.03121	0.013347	0.016949	0.013347	2.338273	0.572272	0.572272	0.572272	0.572272
N18	KGMEGIA	3.381661	3.302672	2.962326	3.381631	3.302672	0.435671	3.304359	36.30417	6.770767	78.69666	0	0.596441	13.92913	0.783542	2.516964	2.613233	1.972918	2.047578	2.047578	33.7	34.2	0.036612	0.029544	0.03039	0.029544	1.313696	0.237615	0.237615	0.237615	0.237615
N19	SASANI	0.705046	0.646229	0.646229	0.705046	0.646229	0.05666	4.672625	40.1958	4.775642	10.76558	0	0.955069	2.649184	2.072066	0.23145	0.124364	0.249472	0.257693	0.257693	37.2	40.2	0.197147	0.090669	0.12549	0.090669	2.169568	0.494839	0.494839	0.626347	0.494839
N20	ELLERINE	3.607202	3.501514	3.275169	3.607202	3.501514	0.057911	1.787256	17.2321	32.64454	214.2236	0	0.930517	15.15184	1.414699	0.148672	0.104006	0.13663	0.14714	0.14714	285.4	287	0.058343	0.038375	0.12098	0.038375	1.520337	0.215023	0.215023	0.36595	0.215023
N21	GLOVIL	0.612629	0.564067	0.564067	0.612629	0.564067	0.059695	27.51979	114.0726	16.05016	-6.27631	0	0.917398	1.773969	1.128266	0.06646	0.059456	0.076065	0.067063	0.067063	1.2	2.9	0.02702	0.02197	0.006093	0.02197	1.22966	0.166699	0.166699	0.876609	0.166699
N22	SEARMAN	1.513063	0.570958	0.546966	1.513063	0.570958	0.078061	7.175718	3.116536	7.696905	120.5739	0	1.663199	14.45119	4.316645	0.111662	0.063466	0.314167	0.27496	0.27406	41.73333	43.46667	0.125261	0.046268	0.294166	0.046268	2.595513	0.61472	0.622665	0.76363	0.61472
N23	UCS	4.333257	4.245392	4.245392	4.333257	4.245392	0.529652	4.371906	49.96651	5.811669	27.96345	0	0.717654	11.25969	0.66026	0.212621	0.259376	0.239976	0.223132	0.223132	11.9	12.6	0.002466	0.002076	0.002294	0.002076	1.198377	0.165512	0.570226	0.666267	0.165512
N24	CONNECT	0.816666	0.501753	0.501753	0.816666	0.501753	0.164693	17.17369	15.67254	5.209954	-25.5159	0	4.25746	31.35101	30.97874	0.025667	0.018466	0.395703	0.572117	0.572117	5.6	6.7	0	0	0	0	7.276343	0.662566	0.662566	1.31109	0.662566
N25	COMPAREX	1.342514	1.045362	0.895812	1.342514	1.045362	0.252522	5.192646	6.367944	3.453163	8.103661	0	1.513417	14.01667	4.146107	0.134676	0.06906	0.274669	0.266326	0.266326	149.3	157.6	0.014522	0.005301	0.610643	0.005301	2.739567	0.612532	0.617675	0.746999	0.612532
N26	CITYLDG	1.166769	1.166665	1.166665	1.166769	1.166665	0.041276	7.03229	246.4714	7.306139	3.42635	0	0.336472	0.45672	0.557372	0.459696	0.30073	0.146959	0.167618	0.167618	154.1	140.7	0.52322	0.315655	0.136406	0.315655	1.656517	0.392103	0.441036	0.653972	0.392103
N27	TRNPACO	1.609366	1.152356	1.104478	1.609366	1.152356	0.000265	4.768299	11.51578	8.956058	53.77066	0	1.953306	5.057034	3.863637	0.075799	0.039923	0.179563	0.154256	0.154256	37.1	37.5	0.214036	0.106203	0.181434	0.106203	1.978101	0.479337	0.479337	0.486306	0.479337
N28	ARGENT	1.769028	1.04766	1.84766	1.769028	1.04766	0	4.067482	5.7762	5.020667	79.78615	0	1.143954	2.20611	1.963042	0.092484	0.065949	0.158948	0.12946	0.12946	35.6	37.6	0.19618	0.114323	0.143482	0.114323	1.716015	0.361459	0.403096	0.403096	0.361459
N29	NICTUS	1.436809	0.875023	0.873602	1.436809	0.875023	0.000358	5.41751	6.367464	6.822499	69.51995	0	2.265578	8.467913	5.426343	0.040656	0.012365	0.177067	0.067205	0.067205	12.53333	12.53333	0.204677	0.066294	0.303523	0.066294	2.374167	0.569633	0.569633	0.569633	0.569633
N30	AST GROUP	0.948642	0.611916	0.94148	0.948642	0.911916	0.20365	5.539281	79.67029	8.59664	28.01535	0	1.614994	10.39421	13.7461	0.124256	0.011564	0.05151	0.158954	0.158954	3	26.2	0.965024	0.137225	0.115271	0.137225	7.178145	0.79007	0.807666	1.060062	0.79007
N31	SEKUNJALO	1.108484	0.814004	0.693754	1.108484	0.814004	0.190681	7.639774	8.014364	5.536917	27.39623	0	1.170616	5.559603	24.99913	0.145315	0.060714	0.292249	1.517609	1.517609	19.1	19.1	6.421429	0.300692	0.533953	0.300692	21.35554	0.796699	0.796699	0.820666	0.796699
N32	IMPERIAL	1.29392	0.99659	0.86153	1.29392	0.99659	0.132141	7.056656	6.477413	5.556029	29.09292	0	1.012169	3.613465	2.323713	0.067964	0.057696	0.18199	0.134536	0.134536	262.3624	364.8	0.293396	0.127601	0.33704	0.127601	2.29573	0.532066	0.534525	0.544912	0.532066
N33	MOBILE	1.048899	1.048899	0.045333	1.048899	1.048899	0	9.304144	59	34.43501	34.61659	0	0.040932	1045.737	0.061603	0.58098	-0.72412	0.001408	-0.04461	-0.04461	-3	13.5	0.354469	0.235542	0.197148	0.235542	1.504993	0.332587	0.332587	0.332587	0.332587
N34	MAXTEC	1.487636	0.620298	0.620298	1.487636	0.620298	0	9.136399	8.534106	9.122958	55.80186	0	2.66612	22.65391	5.159077	-0.10454	-0.06561	-0.31562	-0.44166	-0.44166	-11.6	-7.2	0.008669	0.004631	0.115147	0.004631	1.919214	0.478953	0.545118	0.659944	0.478953
N35	PRIME	0.756806	0.692731	0.504515	0.756806	0.692731	0.03658	8.661097	45.03665	8.462404	4.900063	0	1.267458	6.226048	7.395002	0.215345	-0.01764	0.092458	-0.13046	-0.13046	-10.4	9	1.733312	0.297079	0.393336	0.297079	5.834513	0.736276	0.82742	1.256692	0.736276
N36	MUSTEK	2.127116	1.266909	1.17967	2.127116	1.266909	0.031659	5.306781	6.619709	10.48316	67.48368	0	2.406556	105.7366	4.13367	0.064936	0.070079	0.29025	0.269697	0.269697	90.6	86.4	0.011436	0.006657	0.077931	0.006657	1.717752	0.417644	0.692269	0.81961	0.417644
N37	ADCORP	1.22123	1.22123	1.213655	1.22123	1.22123	0.376664	6.203651	59	13.2874	37.61171	0	2.676164	16.62344	9.613143	0.646918	0.077074	0.53229	0.733214	0.733214	152.5	151	0.182044	0.051248	0.036512	0.051248	3.652114	0.717204	0.740646	0.806536	0.717204
N38	GOLDREEF	1.502014	1.502014	1.502014	1.502014	1.502014	0.176057	16.25356	59	15.38462	4.916066	0	0.072257	392.1569	0.062126	0.245	1.895975	0.168083	0.155709	0.155709	40.6	48	0	0	0.084018	0	1.136586	0.120173	0.281279	0.281279	0.120173
N39	HERCOL	1.362995	1.134633	1.027917	1.362995	1.134633	0.036819	2.120061	8.162963	13.63676	190.3906	0	1.13849	10.29366	3.175597	-0.36415	-0.45212	-1.0933	-1.43575	-1.43575	-69.7	-65.7	0.22677	0.0613	0.311621	0.0613	2.769307	0.941466	0.649926	0.669344	0.641466
N40	STEENS	1.370024	0.971606	0.746766	1.370024	0.971606	0.116559	10.62417	12.9521	12.71096	33.16619	0	2.211151	16.90355	4.207993	0.115767	0.066196	0.283517	0.286976	0.286976	22.5	25	0.063644	0.044057	0.199623	0.044057	1.903076	0.472545	0.503056	0.630676	0.472545
N41	SUBBY	2.446219	1.533266	1.52566	2.446218	1.533266	0.236623	4.141247	2.866672	5.55032	149.7011	0	0.699364	7.710961	1.046676	0.447074	0.4529	0.43355	0.475037	0.475037	40.2	40.2	0.096966	0.066001	0.207222	0.066001	1.499759	0.333226	0.364175	0.364175	0.333226
N42	FRONTRNGE	4.079606	4.061093	3.762094	4.079606	4.061093	0.544462	4.32685	142.0621	4.623631	11.25991	0	0.459905	7.761602	0.752261	1.069366	0.782706	0.444534	0.588616	0.588616	280.9	31.2	0.971226	0.043544	0.000966	0.043544	1.635731	0.21652	0.21662	0.227186	0.21652
N43	BRANDCO	3.474342	2.170737	2.170737	3.474342	2.170737	0.205417	4.749299	5.063463	7.591392	100.5736	0	1.755561	73.49121	2.394478	0.392063	0.067152	0.174346	0.160795	8.160795	16.6	16.5	0.00261	0.001913	0.003222	0.001913	1.36394	0.26663	0.26683	0.366346	0.26663
N44	SPESCOM	0.735367	0.597196	0.597196	8.735367	0.597196	0.209581	4.258167	11.16197	5.191956	46.05926	0	1.361396	8.844714	17.4778	0.29657	-0.23624	-1.39476	-4.12693	-4.12693	-36.1	10.4	0.285146	0.022537	0.442606	0.022537	12.65227	0.916666	0.993133	0.993133	0.916666
N45	MNET-SS	1.097036	1.097036	0.400606	1.097036	1.097036	0	3.457578	59	1.506921	-130.143	0	1.000666	6.991469	7.365443	0.124773	0.093122	0.406611	0.665662	0.665662	27.6	27.6	0	0	0.05927	0	7.36053	0.722437	0.722437	1.294769	0.722437
N46	PPC	2.40657	1.709272	1.405615	2.40657	1.709272	0.056346	7.666623	9.394612	7.535583	36.01057	0	0.696286	1.15518	1.034014	0.161399	0.14603	0.12631	0.150997	0.150997	518.6	500.2	0.161266	0.106921	0.116264	0.106921	1.460764	0.215211	0.217214	0.232134	0.215211
N47	ZAPTRONIX	5.622439	5.132075	5.132075	5.822439	5.132075	0.290615	3.610235	7.666764	8.22396	105.997	0	0.440265	1.077634	0.659666	-0.75226	-0.43081	-0.14476	-0.241	-0.241	-6.4	-6.4	0	0	0	0	1.271206	0.061066	0.061066	0.134963	0.061066
N48	WETHLYS	2.947943	1.619309	1.619309	2.947943	1.619309	0.331736	22.10674	4.60661	7.036619	40.53695	0	1.143956	4.946207	1.501602	0.176367	0.149461	0.216973	0.224461	0.224461	21.3	21.3	0.035956	0.027394	0.027394	0.027394	1.312641	0.236177	0.236177	0.33292	0.236177
N49	FOSCHINI	3.78356	2.606202	2.474936	3.78356	2.606202	0.006665	2.814703	5.377052	6.046706	162.1161	-1	1.023172	9.056518	1.632793	0.06509	0.03996	0.066037	0.065246	0.065246	48.6	50.1	0.212447	0.133127	0.13965	0.133127	1.586615	0.294746	0.294746	0.294746	0.294746
N50	COMPCLEAR	4.966471	4.242455	4.242455	4.966471	4.242455	0.246312	6.575113	13.21	11.9009	62.47299	0	0.69191	1.7558	1.039765	0.244694	0.169265	0.166064	0.196816	0.196816	12.6	12.5	0	0	0	0	1.165796	0.090505	0.090505	0.102627	0.090505
N51	PALS	1.676617	0.956252	0.956252	1.676617	0.956252	0.030244	4.641476	8.016274	5.867102	74.05904	0	2.242624	17.26665	5.819764	0.102655	0.043711	0.1612	0.254366	0.254366	32.4	32.4	0.161156	0.069614	0.110046	0.069614	2.594646	0.585666	0.585666	0.585666	0.585666
N52	CARGO	0.511312	0.486514	0.430751	0.51524	0.490251	0.818413	7.606617	118.6377	12.51918	21.6668	0	1.249126	1.61655	2.995135	0.057364	0.01263	0.120296	0.03644	0.03644	18.7	18.3	0.239636	0.099991	0.265972	0.099991	2.398582	0.523659	0.523659	0.814162	0.520626
N53	JASCO	1.246566	0.790416	0.786661	1.246566	0.790416	0.122711	5.45631	7.245513	6.919746	84.52345	0	2.241949	25.75266	7.431921	0.082249	0.073372	0.406769	0.545295	0.545295	46.5	62.3	0.046491	0.014626	0.312656	0.014626	3.314936	0.69294	0.69294	0.69294	0.69294
N54	PINNACLE	1.524035	0.842763	0.842763	1.524035	0.842763	0.005622	5.569556	6.766526	8.539213	63.48491	0	2.099253	12.24228	4.841384	0.62136	0.002969	0.066022	0.013673	0.013673	0.6	0.2	0.19096	0.06637	0.219253	0.06637	2.21097	0.540276	0.540276	8.540276	0.540276
N55	LA-GROUP	3.159167	2.035156	2.035156	3.159167	2.035156	0.304654	5.069796	4.610062	11.06445	110.0302	1	1.564306	34.84767	3.165373	0.543929	0.500146	1.111763	1.563158	1.582942	73	73	0.413605	0.2045	0.0408	0.2045	2.0235	0.505607	0.505607	0.520135	0.505607
N56	PUTCO	0.732627	0.490501	0.076981	0.732627	0.490501	0.026283	143.3143	16.41051	6.622603	-31.1246	0	2.450512	4.169687	6.344406	0.028486	0.028232	0.166144	0.179116	0.179116	67.8	37.4	0	0	0.113606	0	2.569013	0.549276	0.622626	0.786374	0.540276
N57	GROUP-5	1.039399	0.797103	0.718009	1.039399	0.797103	0.11968	4.892054	11.02463	3.946602	15.26478	0	1.812466	7.431449	5.911499	0.032337	0.02317	0.241203	0.136971	0.136971	64.2	84.2	0.015023	0.004606	0.08639	0.004606	3.261592	0.663107	0.683107	0.748332	0.669107
N58	CONCOR	1.036937	0.611646	0.247629	1.036937	0.611646	0.073901	18.93629	6.36904	3.439029	-29.7304	0	1.94994	17.95597	7.728397	-0.04868	-0.04994	-0.19737	-0.36594	-0.36594	-404.4	-404.4	0.106755	0.026935	0.174204	0.026935	3.963402	0.741207	1.010035	1.021515	0.741207
N59	CONTROL	1.256531	0.725933	0.586361	1.256531	0.725933	0.072373	8.91086	5.54996	8.617167	63.42363	0	1.330093	5.691478	3.107664	0.073139	0.066926	0.192758	0.27016	0.27016	18.2	8.8	0.253515	0.106496	0.24592	0.106496	2.336577	0.560172	0.564569	0.913245	0.560172
N60	DCENTRIX	1.211164	1.14566	1.132696	1.211164	1.14566	0.273613	5.677429	60.93469	5.66744	6.095622	0	2.391296	13.80259	10.25674	0.052722	0.044648	0.205653	0.457939	0.457939	9.7	11.3	0.562606	0.131215	0.001611	0.131215	4.269196	0.73197	0.73197	0.771636	0.73197
N61	ENSERVE	1.036096	0.994957	0.853246	1.036096	0.994957	0.029022	4.858979	84.47606	6.411511	36.04649	0	1.024166	1.556422	2.937031	0.116331	0.066564	0.154037	0.254299	0.254299	30.1	30.6	0.326233	0.114457	0.156435	0.114457	2.667729	0.395501	0.418166	0.693357	0.395501
N62	SOFTLINE	1.151061	1.120192	1.120192	1.151061	1.120192	0.217092	4.283578	63.06204	3.933691	-1.7862	0	0.635127	21.32753	1.507006	0.840399	0.768613	0.739169	1.156304	1.156304	36.4	36.3	0.064717	0.035704	0.144153	0.035704	2.372761	0.361664	0.361664	0.36923	0.361664
N63	MONEYWB	3.527164	3.527194	3.492686	3.527184	3.527194	0.650784	5.405636	59	15.47756	50.12603	0	1.328906	17.10781	1.665926	1.052731	0.066918	0.021662	0.162184	0.162184	-14.5	1.6	0	0	0	0	1.404109	0.256714	0.256714	0.261933	0.256714
N64	CHEMSERVE	1.65242	0.697651	0.622306	1.66242	0.697651	0.073652	8.447218	5.42223	7.201554	73.24554	0	1.631005	9.177053	4.356373	0.112565	0.070254	0.234262	0.306055	0.306055	130.3	149.2	0.075753	0.033578	0.062596	0.033578	2.256013	0.487377	0.487377	0.592647	0.487377
N65	DATATEC	1.441611	1.066065	1.032451	1.441611	1.066065	0.246602	4.60615	6.579457	2.666915	-0.62224	0	1.205603	17.05204	4.429643	0.041263	0.024118	0.056669	0.106637	0.106637	45.41536	45.41536	0.773217	0.210436	0.067245	0.210435	3.674361	0.702022	0.702022	1.515266	0.702022
N66	WINHOLD	1.265204	6.613218	0.813218	1.265204	0.613218	0.027361	5.126299	8.660217	5.054681	41.11301	1	2.143577	7.232657	8.684601	0.044202	0.015077	0.150526	0.146015	0.146204	10.8	7.6	0.32102	0.071054	0.165922	0.071054	4.517963	0.619681	0.619681	0.916691	0.616691
N67	PASDEC	1.057633	0.547192	0.547192	1.037633	0.547192	0	5.160614	5.744277	5.206754	94.3325	0	1.207625	2.633678	2.321159	-0.07007	-0.20216	-0.3197	-0.46924	-0.46924	-32.82	-32.82	0.096505	0.051249	0.247667	0.051249	1.922066	0.479732	0.479732	0.479732	0.479732
N68	OMNIA	1.239164	0.669172	0.669172	1.239164	0.669172	0.039606	5.181274	5.419567	3.956567	45.54263	0	1.676684	5.319946	4.167304	0.111136	0.066562	0.402067	0.28572	0.28572	186.6	171.8	0.128161	0.051563	0.091995	0.051563	2.496148	0.59442	0.634717	0.841753	0.59442
N69	ERP.COM	2.105649	2.076779	2.07323	2.105649	2.076779	0.544692	6.306025	173.51	6.06122	-0.23412	0	2.134307	27.69166	3.716777	0.15293	0.111652	0.403636	0.414675	0.414675	6.6	7.2	0	0	0	0	1.742361	0.426073	0.434205	0.465574	0.426073
N70	MTN	1.049961	0.896661	9.691694	1.049961	0.896661	0.063559	5.135912	21.56429	3.460014	-18.6903	0	0.846847	1.24357	2.008378	1.067476	0.131059	0.19432	0.263216	0.263216	25.1	25.1	0.649644	0.274011	0.241663	0.274011	2.371595	0.530148	0.768523	1.166379	0.530148
N71	MATHOMO	0.720675	0.381666	0.381609	0.720675	0.381609	0	7.426282	8.384669	9.461596	54.19558	0	2.863734	15.7486	-37.7222	-0.13977	-0.19545	42.96656	7.37266	7.37266	-71.7	-39.3	-0.672	0.036199	0.773247	0.086199	-13.1724	1.075916	1.132126	1.144131	1.075916
N72	MASONITE	1.929701	1.11802	0.953827	1.864927	1.070397	0.033195	8.485274	8.6655	9.660156	73.25667	0	1.50178	3.241648	2.470494	0.010114	0.003531	0.021912	0.008722	8.008722	13	12	0.00304	0.001846	0.090506	0.001846	1.645044	0.279428	0.279428	0.324345	0.262496
N73	ADONIS	5.942261	2.269497	2.200395	5.942251	2.269497	0.153071	5.858674	1.660419	6.171767	220.3895	0	0.823933	4.130623	1.308563	-0.01719	-0.03855	0.014677	-0.05037	-0.05037	-13.3	-13.3	0.344514	0.217251	0.216263	0.216263	1.585789	0.350751	0.350751	0.350751	0.349763
N74	APLITEC	2.090411	1.923656	1.923656	2.090411	1.923656	0.215531	3.236479	26.50206	15.30655	162.7065	0	1.698646	11.22801	2.967261	0.161556	0.136495	0.218634	0.407745	0.407745	30.3	34.9	0.035059	0.019936	0.036092	0.019936	1.758406	0.404609	0.404609	0.60734	0.404609
N75	GRINTEK	1.324555	0.809532	0.773254	1.324555	0.809532	0.151622	4.42941	4.707538	2.61741	20.46813	0	1.629504	17.83812	7.493662	0.059119	0.041667	0.276028	0.313737	0.313737	19	16.4	0.352471	0.076645	0.0211	0.076645	4.566737	0.748746	0.750968	0.760489	0.748746
N76	DELTA	1.462075	0.692431	0.692431	1.462076	0.692431	0.136163	6.597626	2.621149	5.417265	112.6949	0	0.929627	2.526402	1.604189	0.267277	0.196672	0.344194	0.315499	0.315499	448.2	446	0	0	0.121621	0	1.725626	0.403017	0.403017	0.669995	0.403017
N77	AMAPS	1.603747	0.661362	0.627172	1.603747	0.661362	0.021329	3.295339	3.469925	4.366108	132.132	0	1.34875	16.50748	3.858156	0.050493	0.054484	0.153262	0.210206	0.210206	12.6	12.5	0.371368	0.129624	0.046006	0.129624	2.860542	0.650416	0.650416	0.669796	0.650416
N78	STELLA	4.611157	1.054962	0.660767	4.611167	1.054962	0.130479	33.06986	1.277114	45.76796	266.6629	0	0.797113	25.50865	1.507627	0.105062	0.063567	0.066071	0.06077	0.06077	1.3	1.3	0.56961	0.295836	0.352995	0.295836	1.691609	0.47136	0.47136	0.50777	0.47136
N79	VALUE	1.144296	1.015606	0.979723	1.144296	1.015606	0.001021	3.746759	29.41973	4.646084	31.73372	0	1.03937	1.565576	2.551798	0.100621	0.056426	0.113327	0.149091	0.149091	10.2	10.7	0.596687	0.243116	0.272399	0.243116	2.455137	0.527306	0.563556	1.074022	0.527306
N80	REBSERV	1.802102	1.37852	1.37852	1.802102	1.37852	0.496229	19.79225	13.44071	8.458432	-10.9174	0	2.600257	26.56661	8.296142	0.372204	0.251202	1.206163	2.084004	2.084004	267.7	110.2	0.606925	0.190855	0.005975	0.190855	3.190506	0.947582	0.66158	0.773037	0.647582
N81	MT-EAGLE	1.000093	0.520211	0.454107	1.000093	0.520211	0.029973	7.367122	4.839638	5.40706	57.52345	0	0.602921	1.157999	1.018333	0.133647	0.076604	0.069461	0.076008	0.076008	29.3	6.9	0.157581	0.077614	0.162996	0.077614	2.024637	0.294353	0.305394	0.3069	0.294353
N82	AF-&-OVER	4.667891	2.264431	2.277692	4.667891	2.264431	0.281173	11.96116	3.346576	14.1847	113.761	1	1.470309	9.176779	3.249966	0.049665	0.026371	0.047484	0.065705	0.063974	67	67	0.006163	0.002797	0.615857	0.002797	2.210396	0.171552	0.171552	0.203567	0.171552
	Max	5.942251	5.132075	5.132075	5.942251	5.132075	0.650784	143.3143	248.4714	45.76796	458.6893	1	4.25746	1045.737	925.9536	2.516964	2.613233	42.96656	7.37266	7.37266	516.9	500.2	254.2093	0.554057	0.773247	0.554057	458.814	1.075916	1.132126	1.515266	1.075916
	Average	1.971023	1.394635	1.306585	1.967776	1.390648	0.151428	8.256029	26.15041	9.037075	60.16061	0.036585	1.524323	31.00093	14.73606	0.176839	0.107566	0.891362	-3.29628	-3.29628	63.93269	51.65693	3.318926	0.082775	0.142648	0.082763	6.015977	0.486854	0.517652	0.651905	0.486971

All Year 1 Potential Variables Selected (Format: Variable Name (Year Prior to Failure))																													Short Name	Ref Code
C1(1)	C2(1)	C3(1)	C4(1)	C5(1)	C6(1)	C7(1)	C8(1)	C9(1)	C10(1)	C11(1)	C12(1)	M1(1)	M2(1)	M3(1)	M4(1)	M5(1)	M6(1)	M7(1)	M8(1)	G1(1)	G2(1)	G3(1)	G4(1)	G5(1)	G6(1)	N1(1)	N2(1)	N3(1)		
0.007578	0.040435	0.009456	426.815	0	0.639717	0.757027	1.942005	0.283193	-0.00087	0.316066	0	0.059922	0.03703	0.066636	4.701681	1.325852	7.821308	-0.31377	-0.38676	5.378758	6.302024	0.382597	1	0.120653	0.035235	0.505564	1	1	CAXTON	N1
0.132572	0.781971	0.036989	420	0	0.908851	0.741975	350	0.27839	0.001973	1.444635	0	0.310754	0.022936	-0.00385	10.36953	3.395886	41.90403	-0.29168	-0.23434	6.048519	6.398966	0.586843	2.66437	-0.03307	0.520009	0.013734	1	1	TRUWTHS	N2
0.049915	0.121676	0.255551	15.13674	0.876905	0.069776	0.094675	-1.23229	-0.93145	0.001232	1.457423	0	0.475367	0.111111	0.277776	1.197633	0.619736	5.494556	-0.60891	-0.60591	5.600415	5.194195	0.158371	0.58456	0.099382	0.077889	0.395135	0	0	INVICTA	N3
-0.65733	-26.757	0.390116	1.032321	0.906651	0.240204	0.135305	0.19778	0.24842	0.118402	2.334647	0	0.883479	0	-0.6	0.533799	-6.10589	-15.9367	-0.66906	-1.60092	4.9976	4.748199	-0.11665	1.002701	3.455569	0.157625	0.38612	1	1	AME	N4
0.658122	0.113808	0.14971	0.649713	0.976775	0.203112	0.17265	-0.88525	-0.43217	0.010885	1.015768	0	0.554688	0	0.058849	3.984553	4.377983	8.578219	-0.32612	-0.58305	5.754709	6.971685	0.399935	1	0.25294	0.761146	0.246193	1	1	MGX	N5
0.019649	0.58326	0.039909	96.91537	0.012213	0.184315	0.19185	5.006199	0.071595	0.016897	0.445805	0.561561	0.144295	0.019405	0.066411	3.975367	3.210682	10.58108	0.062185	0.122905	6.006215	6.17929	0.290061	0.721467	0.153137	0.571681	0.13	1	1	MRPRICE	N6
0.686902	0.754132	-18.6686	420	0	-0.01475	-0.16128	350	-0.28507	0.006643	0.153503	0	0.062995	0.177215	0.066203	0.555013	0.503729	12.21322	-0.22333	-0.63256	5.318712	4.715384	0.164626	-0.48165	-0.02166	0.050246	0.004362	1	1	HOWDEN	N7
0.002451	0.282197	0.001497	7.131973	0	0.793884	0.941071	0.482206	0.526501	0	1.623129	0	0.062669	0	0.686669	3.833033	2.427261	7.662069	0.030371	-1.49119	4.050773	4.199553	-0.22274	1	0.221316	0.453639	0.325309	1	1	INFOWAVE	N8
0.368765	0.424947	0.874982	0.980031	2.036142	0.142626	0.003635	0.685639	-0.02952	0	0.362206	0	0.356565	0.071429	0.41039	0.364639	0.364771	4.943233	-0.15793	-0.09615	5.974402	5.217446	0.394	1	0.201711	0.060559	0.241996	0	0	BEARDEL	N9
0.059132	0.059132	0.945366	85.65132	0.90756	0.543975	0.476239	16.47253	0.363534	-0.01439	2.519489	0	0.117177	0	0.243373	1.312944	2.878233	5.326307	-0.61594	-0.9619	4.685676	4.623603	0.400056	1	0.672844	1.657673	0.570529	1	1	EOH	N10
0.176576	0.224838	0.218942	0.066811	0.124696	0.318365	0.095667	1.181197	0.41338	-0.00342	0.012952	0	0.299147	0.828571	0.164286	1.86154	0.579782	6.946143	-0.36441	-0.43794	5.570077	5.314997	0.32005	0.826393	0.118901	0	0.469002	1	1	NUWORLD	N11
1.803543	1.41769	1.772152	0.463636	10.12781	0.090698	-0.07228	0	-0.08243	0	0.182437	0	0.097051	0	-0.06389	0.902998	1.305534	19.07394	1.921648	1.897526	4.675484	4.313702	0.047071	1	-0.06365	0.676619	0.136989	0	0	LABAT	N12
-0.16264	-0.15264	-0.25277	-0.57799	-1.6422	-0.32094	-0.39297	0.22514	-0.17769	0	-0.36947	0	0.563953	0	-3.02857	0.223626	102.3791	-3.82426	-0.60335	-1.11015	4.295105	3.64366	-0.45256	1	-298.837	0	0.83398	1	1	AWETHU	N13
0.150194	0.556157	0.187421	1.066771	1.048557	0.273375	0.125943	0.363198	-0.04049	0.010927	0.420222	0.466151	0.361334	0.821667	0.057174	2.823368	2.905643	10.7084	-0.25562	-0.14552	6.279545	6.435475	0.214218	0.66626	0.125979	0.358256	0.022999	1	1	NUCLICKS	N14
0.015393	0.63059	0.014968	13.6	0	0.134373	-0.10724	0	-0.39663	0	0.395653	0	0.053968	0.945956	0.045221	1.961236	0.34506	0.966255	-0.66135	-0.89644	5.281133	4.749321	0.706466	-0.91535	-0.00024	0.054486	0.575429	1	1	REX-TRUE	N15
0.576247	0.576247	0.573226	1.581537	2.831577	0.157241	-0.01456	-0.31122	-0.07441	0	0.234218	0	0.015759	0.020636	0.023512	1.119542	2.860642	9.544354	-0.09628	0.944003	4.381025	4.282214	0.291571	0.24	0.022261	0	0.671228	0	0	SPANJAARD	N16
0.077933	0.231066	0.092661	0.674066	0.146266	0.223302	0.170647	4.743212	0.384359	0	0.130717	0	0.046319	0.179732	0.978049	0.286212	0.382986	4.176486	-0.29282	-0.53543	4.984856	4.199146	-0.03713	0.767296	0.215636	0	0.118096	1	1	INMINS	N17
0.002445	0.009271	0.022633	5.643502	0.182268	0.701695	0.77469	-1.67505	1.547099	0	1.521193	0	0.126379	0	0.147007	5.031154	1.579482	0.494395	-0.13227	-0.8269	5.386558	5.484099	0.278972	1	2.047576	0	0.113226	1	1	KGMEDIA	N18
0.126055	0.212979	-40.1217	-0.00821	-165.296	-0.00151	-0.07952	-0.07053	-0.73443	0.011577	0.469906	0	4.843466	0	0.258333	0.620539	0.686202	2.129417	0.962312	0.157136	5.46919	4.976423	0.097661	1	0.297693	0.343948	0.285352	0	0	SABANI	N19
0.111712	0.33915	0.183867	2.760262	1.14212	0.492626	0.260569	-0.02586	0.147803	0	0.252321	0	0.383524	0.948509	0.19336	2.327452	0.76099	4.59376	-0.38111	-0.62448	6.337007	6.037172	0.617672	0.746155	0.118231	0	0.002195	1	1	ELLERINE	N20
0.223827	0.758346	0.276909	2.58545	0.142474	0.303923	0.269445	0.996536	0.623956	0	0.949451	0	0.256674	0	0.133333	1.146364	0.263963	2.369311	-0.45177	-1.32626	4.595664	3.92737	-0.10315	1	0.667083	0	0.530851	1	1	GLOVIL	N21
0.175703	0.315157	0.219423	3.066461	1.974695	0.242351	0.066596	0.118609	-0.16683	0	1.256029	0	0.014973	0.038647	0.29161	0.845762	1.349424	5.386051	0.407639	1.114416	5.623878	5.339803	0.256299	0.609366	0.221615	0.144769	0.190646	1	1	BEARMAN	N22
0.095005	0.193672	0.096949	69.06667	0.011155	1.242475	1.050907	-116.877	-0.54473	0.065351	6.273732	0	0.194992	0	0.978608	13.24212	2.826914	12.58606	-0.74193	-0.72246	5.273676	5.914662	0.398322	1	0.223132	0	0.391659	1	1	UCS	N23
0.197617	0.571156	0.132803	420	0	0.218069	0.207662	350	0.078251	0	1.177479	0	0.356668	0	0.207407	0.418603	2.627283	7.682356	-0.58606	-1.00012	5.243676	4.801273	0.074022	1	0.572117	0.428571	0.090539	1	1	CONNECT	N24
0.016526	0.11139	0.020225	31.46374	0.066011	0.272294	0.273515	0.365211	-0.0122	0	2.971699	0	0.369089	0.013904	0.039539	4.02212	6.749411	6.621968	-0.19736	0.381785	6.705477	7.097991	0.129769	0.655936	0.197613	0.016532	0.066066	0	0	COMPAREX	N25
0.219878	0.236234	0.223515	0.51579	0.549574	0.415491	0.249724	-0.12372	0.011567	0.000125	0.125332	0.357379	0.09806	0.039301	0.112154	2.541689	1.821186	7.484315	-0.2806	-0.50129	5.625993	5.588974	0.248586	0.606658	0.10197	0.260035	0.182474	0	0	CITYLDG	N26
0.234285	0.354326	0.172641	1.979425	0.942865	0.423573	0.309026	0.371303	0.34836	0	0.044851	0		0	0.374747	0.436505	0.411036	3.317455	-0.19766	-0.9499	5.156682	4.474317	0.465476	1	0.154256	0.016049	0.388435	1	1	TRNPACO	N27
0.339146	0.339146	0.301883	1.18926	1.05531	0.358427	0.169592	0.023182	-0.08866	0	0.211206	0	0.538577	0.094466	0.260315	0.701016	0.456876	4.517226	5.747993	2.663613	5.271142	4.698319	0.404742	0.867466	0.06641	0	0.034825	1	1	ARGENT	N28
0.87691	0.817414	-4.12956	-0.16224	-13.3006	-0.02761	-0.1553	0.378318	-0.10483	0.052702	0.279293	0	0.010856	0.824242	0.07596	0.711739	0.962556	5.302061	-0.25304	1.822389	4.524915	4.132831	0.241684	0.676301	0.945451	0	0.605055	0	0	NICTUS	N29
0.050406	0.097592	0.058391	1.592451	0.527496	0.276569	0.259793	-0.75645	-0.36534	-0.00831	3.319385	0	0.264355	0	0.5125	2.12973	12.97621	5.451416	-0.39636	-0.10071	5.997627	6.123817	0.18463	1	0.158954	0.078976	0.150239	1	1	AST GROUP	N30
0.459716	0.513025	0.399191	0.698613	2.541625	0.293671	0.137467	-0.06206	0.107822	0	0.335797	0	0.027523	0	0.868182	0.062855	1.409694	3.943974	-0.37664	-0.77809	5.470676	4.290292	-0.35921	1	1.517809	0	0.083632	1	1	SEKUNJALO	N31
0.344129	0.386234	0.502091	0.776245	3.3974	0.196445	0.150256	0.057618	0.064266	0.00016	0.514429	0	0.425584	0	0.945431	2.057756	2.513612	10.69098	-1.94494	-1.01197	7.002707	7.942085	0.165122	1	0.134536	0.04016	0.051525	0	0	IMPERIAL	N32
0.909536	0.909563	2.708097	0.94866	17.20194	0.034461	-0.86367	-0.09256	-0.07123	0	0.957885	0	0.250701	0	-0.125	0.667361	0.334052	23.29701	-0.87514	-1.03766	5.957141	5.33342	0.013096	1	-0.04461	0	0.272957	1	1	MOBILE	N33
-0.10214	-0.49359	-0.58193	-10.9901	-2.26221	-0.10627	-0.25637	-1.07882	-0.52199	0	0.456066	0	0.043766	0	-1.05455	0.461916	0.424662	-2.83721	-0.94339	-1.21412	4.641751	3.98664	0.099996	1	-0.44166	0	0.473575	1	1	MAXTEC	N34
0.263577	0.416709	0.750614	0.338262	3.814156	0.136485	0.039754	-0.06721	-0.09783	0.020505	0.574752	0	0.606216	0	-0.02543	0.926039	3.548505	2.611964	-0.41115	-0.61391	6.138672	5.922913	-0.09782	1	-0.13048	0.697276	0.307697	1	1	PRIME	N35
0.144248	0.221467	-0.28449	-11.406	-1.01806	-0.18319	-0.2728	1.194147	-0.02091	0.029516	0.756956	0	0.623603	0	0.095873	4.239622	3.042926	13.43631	-0.09601	-0.37262	5.727414	6.975744	0.43604	1	0.289697	0	0.247862	0	0	MUSTEK	N36
0.012861	0.017583	0.014036	28.96218	0.027871	1.926647	1.616369	4.320605	0.790504	-1.01297	0.296311	0	0.347784	0.009532	0.083199	4.59112	11.69629	1.464973	-0.34861	-0.26176	5.408145	5.925706	0.358121	0.651046	0.624	0	0.087655	1	1	ADCORP	N37
0.084861	0.094861	-0.06425	420	-0.37776	-1.41016	-0.61107	350	0.400791	0.180472	199.0794	0	0.004713	0	0.28	3.663601	0.52772	13.66332	0.430136	-0.66854	5.743183	5.409964	0.199224	1	0.155709	0	0.794156	1	1	GOLDREEF	N38
-0.07789	-0.93732	1.667142	9.22086	17.35476	0.027991	-0.91484	0.148636	-0.21795	0	0.068984	0	0.142063	0	-6.97	0.115114	0.205974	-1.73822	-1.19234	-1.21241	5.080648	3.948965	-0.06602	1	-1.43575	0.087763	0.837812	1	1	HERCOL	N39
0.114261	0.131346	0.114215	0.384169	0.709723	0.59522	0.28492	1.274677	0.270562	0.000662	0.244725	0	0.101626	0.119266	0.206422	1.358384	1.221564	3.479209	-0.41761	-0.25027	5.027435	4.834895	0.299676	0.436467	0.126256	0	0.543002	1	1	STEERS	N40
0	0.003679	0	0.950231	3.304123	0.186209	0.115903	0.651351	-0.14456	0.005466	1.672103	0	0.301585	0	0.190521	1.59469	0.798961	1.600967	-0.31867	-0.83714	5.475599	5.291914	0.198669	0.809404	0.394497	0	0.323593	1	1	BUSBY	N41
0.015015	0.019464	0.021724	6.107006	0.002793	1.630363	1.446696	-2.56371	1.735197	0.254912	3.333482	0	0.740806	0	0.237648	7.263429	2.57246	2.687816	-0.43603	-0.29673	6.67991	6.267551	0.394243	1	0.566816	0	0.265558	1	1	FRONTRNGE	N42
0.014561	0.048792	0.053153	100.9042	0.016674	0.724168	0.659612	2.561265	0.501779	-0.03636	9.325313	0	0.206611	0	0.332	1.212731	0.441381	0.468966	0.026111	-0.96216	5.42233	4.932329	0.266395	0.874559	0.140624	1.266424	0.197408	1	1	BRANDCO	N43
0.053822	0.141823	0.877185	1.159053	16.95156	0.020497	-0.01947	-0.89675	-0.64462	0	2.462014	0	0.82313	0	-0.19641	0.738162	6.561129	1.609405	-0.66292	-1.40606	5.426279	5.256641	0.117912	1	-4.12693	0	0.077173	1	1	SPESCOM	N44
0.144592	0.144996	0.061618	420	0.165087	0.443256	0.294577	350	0.414073	0	1.994459	0	0.048913	0.040625	0.06625	1.466923	7.969915	11.76116	-0.26937	-0.33676	5.622922	5.956921	-0.05666	0.526527	0.362507	3.699924	0.019001	0	0	MNET-SS	N45
0.140592	0.146993	0.993106	1.619755	0.586569	0.821	0.477468	-0.01406	0.131545	0.000353	0.143813	0	0.157601	0.661168	0.100736	4.696767	1.496775	6.979033	-0.13417	-0.21044	6.405995	6.410661	0.413146	0.446011	0.667346	0.019912	0.00054	1	1	PPC	N46
-0.01066	-0.10943	-0.0149	420	0	-2.66596	-2.48753	350	-0.56492	0.559352	0.466405	0	0.094617	0	-0.42667	7.316612	0.754167	-3.1943	-0.62521	-1.10562	4.663623	4.336675	0.212116	1	-0.241	0.131996	0.417417	1	1	ZAPTRONIX	N47
0.00743	0.11265	0.007053	5.410536	0.106264	1.062344	0.643294	-0.50706	0.293581	0	0.60503	0	0.333452	0.016129	0.06671	10.45031	3.287192	12.78103	0.076364	0.06074	5.196726	5.692755	0.263797	0.765224	0.171763	0.100742	0.532047	1	1	WETHLYS	N48
0.22266	0.673622	8.149505	0.750129	1.396419	0.338807	0.097251	-0.43373	-0.3352	0	1.090654	0.479744	0.667328	0.064156	0.0926	1.476152	0.694327	6.46206	-0.73464	-1.00342	0.46434	6.102922	0.459732	0.638959	0.941668	0	0.0018	1	1	FOSCHINI	N49
0.001615	0.020709	0.001235	420	0	3.569685	2.482197	350	-1.35944	-0.1652	0.42671	0	0.146013	0.02037	0.116667	15.31996	1.819468	4.513934	-0.29943	-1.94737	4.513004	4.654932	0.404002	0.619651	0.16136	0	0.451996	1	1	COMPCLEAR	N50
0.234856	0.335488	0.24661	3.151319	0.500206	0.375509	0.197517	-0.32901	0.006903	0	0.481074	0	0.2061	0.09569	0.443636	0.376075	0.571742	2.667676	-0.54012	0.246797	4.520235	3.663323	0.25043	0.612764	0.206761	0	0.5604	0	0	PALS	N51
0.696971	0.696971	0.236464	2.256396	1.18042	0.438118	0.314683	-0.31583	0.166376	0	0.262613	0	0.01725	0.03629	0.150606	0.202656	0.254695	3.910256	-0.56675	-0.53221	5.366945	4.394452	0.385139	0.759356	0.02916	0.11761	0.67755	1	1	CARGO	N52
0.322712	0.402273	1.107627	5.076566	4.211615	0.107209	5.000516	6.694367	0.012569	0.066601	2.366766	0	0.709133	0.021127	0.065493	2.65927	6.567597	6.572	1.591042	2.217446	6.215931	5.612662	0.271637	0.66489	0.373467	0	0.266632	0	0	JASCO	N53
0.926369	0.945607	0.995772	0.465947	5.223913	0.077664	0.000749	-0.15656	-0.26549	0.006609	0.039663	0	0.184314	0	0.15	0.077613	0.09295	10.14629	-0.68135	-1.94737	5.146659	3.772527	0.036676	1	0.013673	0	0.261606	1	1	PINNACLE	N54
0.005692	0.034407	0.023449	1.658182	0.182171	0.419745	0.354187	0.941707	0.295611	0	0.436756	1	0.296602	0.947904	0.145709	3.119666	3.193182	1.464973	-0.26261	-0.03201	5.31362	6.511741	0.326001	0.640565	1.014116	0.319195	0.283119	0	0	LA-GROUP	N55
0.06053	0.06053	0.03405	420	0.679584	0.304668	0.150258	350	0.869394	0	0.220473	0	0.183122	0.293252	0.27561	0.4569	5.64975	5.136809	0.313259	0.007663	5.413336	4.61296	0.250947	0.29073	0.052074	0.278623	0.111465	0	0	PUTCO	N56
0.389426	0.437369	0.401191	17.13354	1.991507	0.115664	0.058629	2.156197	-0.0525	0	0.253596	0	0.377176	0.132597	0.465193	0.132162	0.294459	7.197612	-0.39006	-0.75721	6.166691	5.124397	0.086472	0.657259	0.090025	0.267439	0.004676	0	0	GROUP-5	N57
-0.65996	-0.74199	1.398535	1.86795	5.944711	0.039536	0.006634	1.72196	-0.12054	0	1.186411	0	0.094696	0	-1.0871	0.120667	0.355071	-20.8661	-0.83424	-1.00064	5.743712	4.695959	0.147096	1	-0.36594	0.105715	0.00497	1	1	CONCOR	N58
0.242306	0.30994	-0.40806	-0.59686	-3.79471	-0.116	-0.23177	-0.26067	-0.22859	0	1.021976	0	0.229144	0	0.17234	0.991636	1.256672	7.729572	0.169495	-0.17649	5.18667	4.920002	0.259605	1	0.27016	0.113621	0.332632	1	1	CONTROL	N59
0.0174	0.111555	0.01945	5.946353	0.012166	0.189646	0.093001	-0.65297	0.002644	0	1.079695	0	0.103226	0	0.107776	1.207466	3.790916	12.65657	-0.76651	-1.20217	6.216933	6.163302	0.276066	1	0.457939	0	0.206124	1	1	DCENTRIX	N60
0.166234	0.271677	0.116004	1.729462	0.600381	0.500504	0.300366	-0.61723	-0.45294	0	0.437517	0	0	0.082645	0.24676	0.875049	0.992472	3.5741	-0.19766	0.496636	5.567059	5.106239	0.191206	1	0.254299	0.415173	0.291997	1	1	ENSERVE	N61
0.017727	0.017727	0.067482	4.121456	0.979313	0.406652	0.22391	0.911023	-0.61316	0.09666	13.50424	0	0.64639	0	0.034275	11.11667	9.545137	4.662121	-0.13016	1.919563	5.977666	6.582394	0.273253	1	1.156304	0.24735	0.119337	1	1	SOFTLINE	N62
0.00395	0.022419	0.02666	420	0	0.779529	0.723265	350	-0.75928	0	5.449914	0	0	0.025641	-0.37179	12.4267	4.460335	1.253745	0.947717	-1.03915	4.016574	4.520484	4.262799	1	0.182194	0	0.520165	1	1	MONEYWB	N63
0.057522	0.153924	0.052173	3.463504	0.290643	0.563183	0.358513	0.029197	0.291252	0.051471	9.718695	0	0.182169	0.035171	0.123659	1.944016	2.137506	5.366044	-0.11322	-0.36388	5.91169	5.666256	0.375245	0.743979	0.227666	0.093765	0.000671	1	1	CHEMSERVE	N64
0.312203	0.416517	0.204716	9.360694	1.146433	0.16812	0.096729	-0.60226	0.643361	0.45053	0.589662	0	0.346275	0.004226	0.019626	3.457539	6.916709	45.78401	1.460411	10.36716	6.550453	6.936571	0.041764	1	0.106637	11.50256	0.24747	0	0	DATATEC	N65
0.358205	0.467211	0.327214	1.556686	1.500075	0.178782	0.119011	0.69732	0.269621	0	0.116373	0	0.030602	0.11675	0.68125	0.070962	0.2061	5.0644	-0.32144	-1.24525	5.380797	4.023269	-0.16576	0.909446	0.132793	0	0.203206	1	1	WINHOLD	N66
-0.60524	-1.05626	0.341949	3.507662	1.37667	0.374741	0.264113	2.019324	0.169144	0.169671	0.37726	0	0.312404	0	-0.66366	0.834629	0.769599	-34.4737	-0.97653	-0.4713	4.725201	4.327692	-0.53552	1	-0.46924	0	0.630655	0	0	PASDEC	N67
0.343321	0.391585	0.224349	5.948472	0.265613	0.516201	0.336093	0.73349	0.368231	0	0.234445	0	0.166836	0.972336	0.16436	1.049255	1.549962	5.560975	-0.20465	0.005592	5.603341	5.596316	0.259794	0.566572	0.181651	0.656642	0.195586	0	0	OMNIA	N68
0.000115	0.000219	7.81E-06	420	0	1.260497	0.912526	350	1.242087	0.006496	0.59491	0	0	0	0.276	1.960367	1.455355	3.303543	0.226636	-0.64597	4.666664	4.61149	0.404791	1	0.414675	0	0.177999	1	1	ERP.COM	N69
0.943233	0.056953	0.185346	0.896257	0.982655	0.464271	0.360912	-0.06943	0.06699	0.856447	1.591943	0	0.197444	0.004264	0.026759	6.929747	6.667014	2.336641	0.205047	0.79666	6.235623	9.794325	-0.73149	0.833912	0.219262	0.462419	0.8611	0	0	MTN	N70
-0.35531	-0.37363	12.93399	0.166094	70.32512	0.018219	-0.13424	1.309605	-0.15386	0	0.151921	0	0.091151	0	-7.17	0.87308	-1.03572	-3.86753	0.270351	-0.85231	4.965258	3.660635	-1.25127	1	7.37286	0	0.168166	1	1	MATHOMO	N71
0.933723	0.73703	0.508094	11.21336	4.366077	0.974151	0.667535	10.72203	-0.0334	0	0.083022	0	0.017694	0	0.0325	0.574341	0.284006	5.918113	-0.22661	-0.71501	5.230279	4.435717	0.571572	1	0.006722	0.037076	0.000587	1	1	MASONITE	N72
-4.44079	-4.44078	-1.8	-0.11374	-8.782	-0.07045	-0.18542	-0.50197	-0.69904	0	0.023466	0	0.614345	0	-0.1622	0.556426	0.310606	15.16545	-1.09136	-0.54773	4.191157	3.473122	0.479244	1	-0.05037	0	0.666621	0	0	ADONIS	N73
0.036625	0.190692	0.942246	16.20636	0.111696	0.706196	0.723902	11.65706	0.464394	0.163974	2.336738	0	0.357071	0	0.156995	5.862106	2.620969	4.329112	-0.48903	-0.50923	5.410169	5.583534	0.356444	1	0.407745	0.592661	0.065786	1	1	APLITEC	N74
0.32461	0.455621	0.119925	3.402134	0.03062	0.348256	0.308366	0.052159	0.706277	0.052622	0.436551	0	0.202495	0.025969	0.107345	5.850254	2.927668	11.48079	0.361562	-0.5057	5.762174	5.566057	0.095029	0.769234	0.241336	0.104221	0.010679	0	0	GRINTEK	N75
0.078615	0.103136	0.092631	420	0.499296	0.806612	0.249669	350	0.039782	0	0.307565	0	0.156124	0.033047	0.667642	5.162394	3.590225	6.260287	0.084707	5.761354	6.062251	6.400429	0.433356	0.706072	0.222765	0.404933	0.009378	1	1	DELTA	N76
0	0.166727	0	-0.11725	-3.0224	-0.0234	-0.0242	-0.92027	-0.94123	0.010351	2.931007	0	0.322771	0.00995	0.062667	1.816524	3.3653	21.16585	0.039071	5.606676	5.215474	5.266623	0.099366	0.634696	0.175494	0.077617	0.467787	1	0	AMAPS	N77
0.290195	0.382334	1.636773	0.056302	20.46563	0.036593	-0.03292	0.036347	-0.69241	0	0.016936	0	0.044066	0	0.1	0.926656	0.626176	9.100409	-0.73382	-0.94076	4.61676	4.260071	0.06822	1	0.06077	0	0.654136	1	1	STELLA	N78
0.194203	0.43561	0.096336	0.870779	1.286711	0.461476	0.293597	-0.66957	-0.15616	0	0.566096	0	0.190463	0	0.299421	0.477349	0.61796	4.232663	-0.63121	-0.9566	5.49911	4.697009	0.275069	1	0.149091	0.037309	0.540003	1	1	VALUE	N79
0.992067	0.028157	0.014243	0.771682	0.040553	0.227512	0.226666	-1.28603	-0.16605	0.046561	4.063677	0	0.363627	0.006743	0.14607	4.019026	6.303768	2.875425	-0.2236	0.930136	6.113152	6.528566	0.313296	0.95915	1.996672	0.036122	0.194253	0	0	REBSERV	N80
0.36894	0.36894	1.635905	0.163953	11.3971	0.048629	-0.10421	-0.06063	0.007093	0	0.166269	0	0.051905	0.006611	0.023478	5.362434	3.196096	25.29447	0.620239	-0.92077	4.667325	4.696565	0.120034	0.743418	0.057993	0.003467	0.604444	1	1	MT-EAGLE	N81
0.006925	0.414567	0.009235	29.97273	0.210179	0.43976	0.208941	-0.96	-0.08342	0	0.209671	0	0.042417	0.039731	0.091007	2.460904	0.933172	5.164715	0.126945	0.131648	5.293645	4.919137	0.397654	0.742213	0.063612	0.19966	1	1	1	AF-&-OVER	N82
1.803543	1.41769	12.93399	426.315	70.32512	3.569685	2.482197	350	1.735197	0.856447	199.0794	1	4.843466	0.203252	0.865182	15.31996	102.3791	45.79401	5.747993	10.36715	7.002707	7.097991	4.262799	2.66437	7.37286	11.50255	1	1	1		
0.103515	-0.12666	-0.32021	89.47043	0.246786	0.331821	0.215154	46.38117	0.031081	0.025742	3.541204	0.03551	0.285867	0.027935	-0.12455	2.666469	3.543696	6.319647	-0.15721	-0.21138	5.408618	5.184726	0.233586	0.853102	-3.33556	0.36091	0.288386	0.719512	0.707317		

Appendix F.5.

All Year 3 Potential Variables Selected [Format: Variable Name (Year Prior to Failure)]

Ref Code	Short Name	L1(2)	L2(2)	L3(2)	L4(2)	L5(2)	L6(2)	L7(2)	L8(2)	L9(2)	L10(2)	L11(2)	E1(2)	E2(2)	E3(2)
N1	CAXTON	3.210212	2.664781	2.629246	3.224341	2.669056	0.269496475	4.851956	10.62057	6.049172	49.25583	1	1.12581	3.47083	1.77009
N2	TRUWTHS	1.933716	1.460447	1.055158	1.9332	1.459912	0.018325569	4.449926	6.446226	14.03539	94.65766	0	1.622066	10.96815	2.601181
N3	INVICTA	4.27946	3.250248	3.250249	4.27946	3.358249	0.51990596	6.657876	6.184506	7.667676	64.76332	0	1.42513	37.70026	1.63634
N4	AME	1.195296	0.917641	0.917641	1.195296	0.917641	0.132193714	5.299541	12.10367	5.586023	33.87618	0	1.667154	14.71367	4.828368
N5	MOX	3.177785	2.798692	2.798692	3.177785	2.798692	0.362119394	4.968069	13.09626	0.14321	41.9669	0	0.996542	3.797621	1.454077
N6	MRPRICE	1.626005	0.705466	0.665342	1.626618	0.704361	0.093498184	11.74469	8.287041	6.536417	33.3099	1	2.691282	12.96116	6.025606
N7	HOWDEN	1.733913	1.226697	1.216904	1.734558	1.224663	0.224210367	4.796026	7.036497	3.916957	34.78646	0	1.667361	11.3655	3.466145
N8	INFOWAVE	1.604104	1.604104	1.534932	1.604104	1.604104	0.240486663	5.410244	59	5.562334	6.266213	0	2.246669	11.48792	4.565006
N9	SEARDEL	1.953916	1.056136	1.056136	1.953916	1.056136	0.095939034	5.256709	4.505071	4.410706	97.87529	0	1.361931	3.924252	8.534654
N10	EOH	2.10712	2.096063	1.966761	2.10712	2.096063	0.512572169	4.666202	340.6711	39.64247	66.83953	0	1.591432	44.0667	3.548364
N11	NUWORLD	2.540304	1.577101	1.577101	2.540304	1.577161	0.159765937	4.180178	4.617639	5.31496	93.26097	0	1.72557	23.14647	3.260155
N12	LABAT	2.226544	2.076641	2.043619	2.226544	2.076641	0.000263261	0.907412	12.36698	22.66766	416.791	0	0.810814	90.78366	2.262411
N13	AWETHU	1.663645	1.236362	1.166409	1.663646	1.236362	0.129499301	8.399986	16.04615	8.300236	9.773723	0	1.164925	3.660066	2.126659
N14	NUCLICKS	1.46194	0.432593	0.432593	1.49164	0.432593	0.119460486	41.34971	5.367427	6.72026	22.26427	0	2.314478	6.131976	4.830525
N15	REX-TRUE	4.930581	2.30286	2.298049	5.046057	2.357612	0.261869196	11.96326	3.346576	14.25562	113.9728	1	1.474416	9.67366	1.776476
N16	SPANJAARS	0.960341	0.71611	0.676301	0.972965	0.725539	0.000234104	5.107293	14.13486	6.076913	37.22572	0	2.184107	6.902776	7.312292
N17	INMINS	1.841294	1.113068	1.113068	1.941294	1.113068	0.000416769	5.876527	6.674827	7.126024	51.55293	0	3.035606	22.46317	7.661227
N18	KGMEDIA	1.93322	1.93322	1.925626	1.93322	1.93322	0.090498262	5.746234	59	7.363063	20.13466	0	0.411421	12.96624	0.479177
N19	SASANI	0.960963	0.663781	0.663791	0.960963	0.663781	0.166271862	3.606183	23.2626	3.314997	1.430953	0	0.720099	2.502627	1.477116
N20	ELLERINE	2.963451	2.666555	2.564917	2.963451	2.666555	0.026466172	1.906713	15.67767	24.67976	199.921	0	0.939971	6.660193	1.466614
N21	GLOVIL	0.401247	0.304127	0.384127	0.401247	0.394127	0.052266206	27.13577	206.4634	9.07166	-44.6946	0	0.726187	1.307644	0.851924
N22	BEARMAN	1.565654	0.566647	0.564632	1.565554	0.566647	0.035262664	6.548371	3.242651	7.29227	118.2413	0	1.633724	9.834565	3.754792
N23	UCS	4.443492	4.411227	4.411227	4.443492	4.411227	0.543917296	5.724769	149.2064	6.161025	26.49627	0	0.715294	10.36169	0.934029
N24	CONNECT	0.631146	0.501461	0.501461	0.631146	0.501451	0.212560996	15.69241	14.46439	4.775412	-27.974	0	5.149751	49.66676	-65.62629
N25	COMPAREX	1.746644	1.456972	1.327163	1.746644	1.456972	0.436700197	3.987293	6.084332	2.903506	25.82084	0	0.663477	26.8113	2.003507
N26	CITYLDG	0.472355	0.452641	0.452641	0.472355	0.452641	0	10.46904	240.3769	6.01329	-24.31574	0	0.335399	0.437009	0.553951
N27	TRNPACO	1.296317	0.95122	0.893433	1.296317	0.95122	0.000276707	4.436604	11.37932	6.660627	61.29466	0	1.691795	3.962677	4.032673
N28	ARGENT	1.777568	1.105699	1.105699	1.777568	1.105699	0.033966293	4.017921	5.909339	4.920297	76.42712	0	1.142791	2.361477	2.032911
N29	NICTUS	1.363377	0.882132	0.799691	1.363377	0.802132	0.001109095	5.657274	7.763551	7.318943	61.66268	0	2.233515	7.621101	5.605736
N30	AST GROUP	1.390266	1.379243	1.345931	1.390266	1.378243	0.302169774	5.43185	260.8166	6.605122	13.39135	0	1.521736	10.87597	2.956313
N31	SEKUNJALO	1.060696	0.667716	0.667716	1.000693	0.667716	0.058693692	4.792064	6.201974	5.056618	02.84824	0	1.127465	4.167374	-299.4734
N32	IMPERIAL	1.444619	1.015076	0.81792	1.444618	1.015076	0.070352661	6.935363	8.32527	6.076661	24.62547	0	1.157151	4.769239	2.247619
N33	MOBILE	0.909378	0.909378	0.007604	0.909378	0.909378	0	65.89605	96	11.90682	-20.63186	0	0.066359	1827.543	0.095256
N34	MAXTEC	1.476449	0.566716	0.566716	1.476449	0.566716	0.052963767	6.051866	3.16157	4.726619	100.7752	0	1.441196	17.56219	3.06306
N35	PRIME	0.770932	0.73165	0.73165	0.770932	0.73165	0.026143223	4.165491	70.69196	5.326601	24.24472	0	1.102706	4.623665	5.171619
N36	MUSTEK	1.787491	0.692266	0.714119	1.787491	0.692266	0.015514795	8.670549	6.566367	6.966969	45.35184	0	2.699639	96.03232	5.362
N37	ADCORP	2.072302	2.072302	2.070934	2.072302	2.072302	0.329150465	6.16302	59	6.260378	21.21517	0	2.152717	8.916447	4.996646
N38	GOLDREEF	1.650667	1.050667	0.586262	1.690667	1.690667	0.072869451	29.44063	66	12.99545	-9.502479	0	0.134318	384.8164	0.18366
N39	HERCOL	2.095184	1.844926	1.716267	2.095184	1.844926	0.014114208	1.530649	10.27336	11.29069	241.6623	0	0.967332	11.996	1.766778
N40	STEERS	1.29146	0.962605	0.960607	1.29146	0.962605	0.147036162	5.152942	19.36627	16.74118	41.27263	0	2.441747	15.56169	6.305036
N41	BUSBY	4.177134	2.677532	2.977532	4.177134	2.977532	0.299660757	5.693625	3.295161	6.510693	110.6104	0	0.61882	10.12731	0.677479
N42	FRONTRNGE	0.667944	0.670403	0.632744	0.667944	0.670403	0.244208579	3.320365	75.46763	3.613432	13.75216	0	0.744769	10.11542	2.361665
N43	BRANDCO	2.643443	1.666184	1.666184	2.643443	1.660184	0.160039777	4.403305	4.454066	7.146766	113.7621	0	1.766563	43.47517	2.677923
N44	SPESCOM	1.769974	1.5171	1.51364	1.769974	1.5171	0.165314134	3.966246	14.6251	6.105066	45.43995	0	1.455378	6.699016	2.450679
N45	MNET-SS	1.061403	1.061403	0.402618	1.061403	1.061403	0	4.366165	59	2.259678	-72.09166	0	1.253362	6.262669	11.13663
N46	PPC	1.413604	0.69497	0.677248	1.413694	6.69467	0.031240702	7.179277	9.479699	8.297664	50.15001	0	0.601646	1.243245	1.263967
N47	ZAPTRONIX	1.302151	1.109916	1.109916	1.302151	1.109916	0.114060929	2.916405	11.62236	10.20737	120.6719	0	0.739421	1.653246	1.606341
N48	WETHLYS	3.042172	1.725007	1.725007	3.042172	1.725007	0.357708854	28.96336	3.812028	8.710188	75.66524	0	1.066236	4.217118	1.413637
N49	FOSCHNI	4.008494	2.899976	2.708218	4.008494	2.899979	0.981739625	2.37909	5.089923	7.801417	176.7857	-1	1.003032	6.265147	1.507718
N50	COMPCLEAR	5.599154	5	5	5.599154	5	0.383646533	6.848213	12.60646	14.85493	59.50921	0	0.75008	1.867414	0.866877
N51	PALS	1.445252	0.796694	0.796694	1.445252	0.796694	0	4.327276	6.315668	4.742143	79.04422	0	2.078774	16.61739	6.407926
N52	CARGO	0.583552	0.557477	0.557477	0.588002	0.662301	0.056775018	6.306781	102.4798	8.849254	7.313326	0	1.122636	1.515497	2.524479
N53	JASCO	1.630796	1.058541	1.046512	1.630796	1.058541	0.198811094	4.378088	6.131794	6.007993	76.68055	0	1.495836	14.30804	4.205117
N54	PINNACLE	1.365129	0.714101	0.714101	1.365129	0.714101	0.002080867	4.941727	5.2919	4.490498	63.20747	0	1.841384	12.09484	4.804817
N55	LA-GROUP	1.655908	1.206507	1.206507	1.655908	1.206507	0.286627924	5.795047	9.130639	11.28066	70.62112	1	2.361417	51.51363	5.549837
N56	PUTCO	0.621473	0.613267	0.106172	0.621473	0.613267	0.012167926	65.62087	24.09852	6.798363	-33.20106	0	2.566673	4.491111	6.666995
N57	GROUP-5	1.015787	0.765211	0.672016	1.015787	0.765211	0.104497835	5.258695	10.97586	3.959332	10.47282	0	1.928239	7.867434	7.098528
N58	CONCOR	1.241177	0.56372	0.549331	1.241177	0.56372	0.184257049	12.6847	4.801271	3.89512	11.13481	0	2.04929	16.60253	6.032355
N59	CONTROL	1.306847	0.847786	0.947786	1.306847	0.947786	0.215814206	4.50271	6.134532	3.844648	45.62437	0	1.036573	5.106615	2.192047
N60	DCENTRIX	1.326237	1.290616	1.280265	1.307266	1.280645	0.371565413	5.121673	122.1421	4.393227	-8.628336	0	1.630579	11.06203	7.582158
N61	ENSERVE	1.007216	0.960639	0.666676	1.007210	0.960639	0.053199319	4.903666	70.60612	21.95117	62.96315	0	1.253434	2.145151	3.907378
N62	SOFTLINE	2.900166	2.649699	2.243722	2.900166	2.649699	0.355356417	7.365739	73.34241	6.261036	-14.76021	0	0.752261	31.62493	1.148077
N63	MONEYWB	2.804999	2.804999	2.786686	2.804999	2.804999	0.577219633	5.066665	59	11.06299	45.25105	0	0.084161	3.71208	0.959379
N64	CHEMSERVE	1.565323	0.900699	0.900699	1.565323	0.900699	0.063514666	5.269384	0.956336	6.362541	72.83637	0	1.993676	8.818015	4.60144
N65	DATATEC	1.569606	0.966977	0.966977	1.569606	0.966977	0.124606051	4.404046	4.960357	3.210262	42.76469	0	1.269303	10.36179	4.840016
N66	WINHOLD	1.361412	0.666634	0.666634	1.361412	0.666634	0.022339423	5.119021	8.19401	5.669764	61.69715	1	1.962336	6.715636	0.73279
N67	PASDEC	1.202647	0.550265	0.550265	1.202647	0.550265	0	5.194061	4.391042	5.552494	67.84966	0	1.405395	3.439262	3.313352
N68	OMNIA	1.144697	0.791546	0.791546	1.144697	0.791546	0.020366201	3.966576	9.006513	4.204241	65.44135	0	1.599906	5.273556	5.119615
N69	ERP.COM	1.757266	1.676196	1.676196	1.757266	1.676196	0.301121112	4.566459	69.63645	6.720043	21.65636	0	2.182663	12.76125	4.064266
N70	MTN	1.296093	1.164718	1.164719	1.296093	1.164716	0.131460042	6.369443	34.16333	4.370344	-15.44477	0	0.665564	1.29263	1.902702
N71	MATHOMO	1.310552	0.936139	0.751571	1.310552	0.936139	0	2.724666	6.470474	5.640521	135.9619	0	1.320711	11.17336	4.752658
N72	MASONITE	1.764265	1.047576	0.831340	1.727843	0.960066	0.021026632	6.659217	6.765963	6.359604	51.36431	0	1.455398	3.613781	2.267163
N73	ADONIS	2.752448	1.164619	1.135445	2.752446	1.104619	0.09436791	4.632481	2.279921	4.066727	151.3632	0	1.014719	4.569437	1.728141
N74	AFLITEC	2.662527	2.541026	2.541023	2.662527	2.541023	0.345600477	2.666121	32.12349	6.297131	66.96062	0	1.142992	5.196262	2.012602
N75	GRINTEK	1.439215	0.99141	0.990713	1.439215	0.99141	0.048203061	11.09417	22.50063	13.12526	21.31266	0	6.667994	43.16732	20.62323
N76	DELTA	1.512459	0.786339	0.786339	1.512459	0.786339	0.116219493	8.143662	4.251694	7.10632	93.66399	0	1.264213	3.666397	2.357692
N77	AMAPS	2.074617	1.160786	1.110656	2.074617	1.160786	0.136419318	11.90117	10.65066	16.73064	45.45256	0	4.179947	46.65669	7.450563
N78	STELLA	2.180663	0.973303	0.654471	2.180663	0.973303	0.116646316	2.367136	0.969705	1.593625	302.2576	0	0.368267	9.096162	1.037376
N79	VALUE	1.093687	1.009963	1.009963	1.093687	1.009963	0.000640462	4.44699	63.5116	6.613694	33.65504	0	1.161176	1.774431	2.196138
N80	REBSERV	1.600177	1.394137	1.393632	1.606177	1.394137	0.35161963	11.16995	10.23016	4.636592	-10.33169	0	1.429195	20.66366	3.727664
N81	MT-EAGLE	0.911952	0.534357	0.479131	0.911952	0.534357	0.039265069	10.26172	8.973604	9.129618	36.19406	0	0.669542	2.271050	-1.631396
N82	AF-&-OVER	4.53873	2.346715	2.344179	4.53873	2.346715	0.250764209	8.02006	3.43094	10.62965	117.5674	1	1.336666	7.729106	3.066094
	Max	5.599154	5	5	5.599154	5	0.577219633	65.89605	346.9711	39.64247	415.791	1	6.667994	1827.543	20.62323
	Average	1.670466	1.375446	1.291262	1.671123	1.375242	0.149906941	6.935967	34.24266	6.003006	66.31133	0.060979	1.516265	39.96337	-0.610709

Ref Code	Short Name	P1(2)	P2(2)	P3(2)	P4(2)	P5(2)	P6(2)	P7(2)	S1(2)	S2(2)	S3(2)	S4(2)	S5(2)	S6(2)	S7(2)	S8(2)
N1	CAXTON	0.130736	0.057426	0.098301	0.11935	0.119318	412	398	0.000439693	0.000279605	0	0.000279605	1.572142	0.196644	0.201545	0.211062
N2	TRUWTHS	0.971769	0.06744	0.185116	0.175422	0.175422	26.6	28.2	0.016619397	0.01161094	0.122432	0.01191094	1.603606	0.37444	0.379115	1.657706
N3	INVICTA	0.110062	0.103224	0.184193	0.189555	0.189555	84	54	0	0	0.001036	0	1.286542	0.223695	0.267073	0.290992
N4	AME	0.164744	0.156183	0.478515	0.754112	0.754112	1.5	2.6	0.584457636	0.169696012	0.286691	0.169696012	3.06099	0.666012	0.661526	0.696618
N5	MOX	0.300056	0.35429	0.33407	0.618165	0.618165	298.6	165.2	0.136020996	0.062721416	0.016439	0.092721418	1.456201	0.293956	0.293956	0.335304
N6	MRPRICE	0.116192	0.096664	0.512642	0.486665	0.486665	30.9	30.6	0.000665693	0.000474406	0.000464	0.000474406	1.667366	0.464486	0.534294	0.636978
N7	HOWDEN	0.014667	0.006619	0.15186	0.024133	0.024133	3.9	3.9	0.002447174	0.001163793	0	0	2.067231	0.473627	0.473627	0.473913
N8	INFOWAVE	0.486032	0.569197	2.247203	2.609727	2.609727	-9.3	-9.1	0.167403537	0.062101321	0	0.062101321	2.036967	0.50956	0.50956	1.06269
N9	SEARDEL	0.04991	0.010902	0.056643	0.971242	0.971242	26.3	26.5	0.32800269	0.067659329	0.092929	0.067659328	4.633571	0.402118	0.410648	0.426536
N10	EOH	0.260511	0.136858	0.336962	0.485623	0.485623	19.1	18.1	0.288496686	0.129390616	0.010466	0.129390618	2.229668	0.551503	0.551503	0.769462
N11	NUWORLD	0.070976	0.044734	0.143651	0.145641	0.145641	134.8	134.8	0.050460271	0.026706156	0.029429	0.026706156	1.669321	0.391615	0.361015	0.443619
N12	LABAT	0.122312	0.031626	0.166104	0.071551	0.071551	22.8	22.8	0.013014646	0.004664218	0.406606	0.004664218	2.790295	0.441981	0.441981	0.68334
N13	AWETHU	0.360104	0.355644	0.609261	0.757163	0.757163	2.7	2.7	0.366963656	0.215374043	0.231653	0.215374043	1.796706	0.443426	0.443426	0.443426
N14	NUCLICKS	0.070945	0.042408	0.165518	0.204846	0.204846	55.8	58.8	0.240692159	0.115324268	0.131724	0.115324268	2.06709	0.520664	0.520664	0.652104
N15	REX-TRUE	0.040774	0.036418	0.066662	0.06625	0.066152	53.8	53.6	0.001720636	0.001427969	0.615902	0.001427969	1.204667	0.166079	0.166079	0.201163
N16	SPANJAARS	0.063317	0.037924	0.364613	0.275115	0.275115	27.6	30.5	0.22923566	0.067643431	0.270999	0.067643431	3.376696	0.694726	0.694726	0.694726
N17	INMINS	0.024	0.01943	0.143045	0.148659	0.148659	12.9	12.9	0.361576363	0.139305023	0.025246	0.139305023	2.523766	0.60377	0.60377	0.60377
N18	KGMEDIA	0.425365	-0.066475	-0.034021	-0.841599	-0.041599	-9.2	31.7	0.029343547	0.025676611	0.036573	0.025676611	1.142612	0.109647	0.109647	0.109647
N19	SASANI	0.439792	0.107606	0.156656	0.156947	0.156947	3.6	4	0.165163692	0.090277541	0.047519	0.090277541	2.051273	0.491241	0.491241	1.279751
N20	ELLERINE	0.153443	0.195769	0.155561	0.161615	0.161915	266.9	280.6	0.062770554	0.039635691	0.142466	0.039635691	1.56366	0.242948	0.243152	0.436527
N21	GLOVIL	-0.311428	-0.27015	-0.229133	-0.257061	-0.257061	-11.7	-1.4	0.037765374	0.026910373	0.0039	0.026910373	1.306963	0.234679	0.234679	0.924366
N22	BEARMAN	0.066326	0.046706	0.23023	0.175377	0.176377	24.2	20.5	0.136160029	0.060549526	0.212489	0.060549526	2.296264	0.564693	0.593236	0.76542
N23	UCS	0.331486	0.336696	0.279669	0.313455	0.313455	12.5	13	0.107330914	0.062196796	0.00046	0.062196796	1.305796	0.233623	0.972248	1.096966
N24	CONNECT	0.046323	0.040669	5.62661	-2.666959	-2.666969	43.2	51.2	0	0	0	0	-12.7245	1.078369	1.078369	1.511027
N25	COMPAREX	0.679462	0.447582	0.796376	0.696734	0.696734	608.6	105.4	0.01719036	0.007580354	0.021572	0.007580354	2.267752	0.505076	0.553905	0.631966
N26	CITYLDG	0.482523	0.304475	0.161327	0.166664	0.166664	139	139	0.526611905	0.316646267	0.119661	0.316646267	1.651617	0.369625	0.435509	0.569042
N27	TRNPACO	0.020477	-0.000461	0.071284	-0.001618	-0.001618	-0.5	-2.4	0.342401184	0.143644843	0.320103	0.143644843	2.363665	0.570749	0.570749	0.573108
N28	ARGENT	0.090367	0.063215	0.169037	0.12851	0.12851	6.7	7.1	0.198666447	0.111660484	0.143036	0.111660484	1.7789	0.399602	0.434515	0.434515
N29	NICTUS	0.943111	0.020224	0.169202	0.11337	0.11337	19.6	18.9	0.2471631	0.098474155	0.279391	0.098474155	2.509627	0.593433	0.593433	0.593433
N30	AST GROUP	0.247575	0.102406	0.277403	0.302751	0.302751	21.9	16.1	0.099631044	0.05138713	0.023566	0.05138713	1.942726	0.48404	0.499561	0.752066
N31	SEKUNJALO	-0.05654	-0.03367	0.075659	10.14302	10.14302	-13.8	0.2	-93.22505543	0.350982127	0.539786	0.350982127	-266.612	0.791249	0.791249	0.791249
N32	IMPERIAL	0.066326	0.056597	0.134721	0.127203	0.127203	270.6	284.6	0.230747665	0.118602218	0.226626	0.118602216	1.942266	0.442234	0.482717	0.490252
N33	MOBILE	0.465434	0.50066	0.847755	0.04771	0.04771	3.6	3.6	0.304749395	0.212299663	0.150952	0.212299663	1.435468	0.300965	0.300965	0.300966
N34	MAXTEC	0.171969	0.096757	0.314233	0.296373	0.296373	8.2	9.2	0.002934758	0.0013606	0.615966	0.0013606	2.125404	0.513268	0.513268	0.569956
N35	PRIME	0.120144	0.02744	0.147367	0.141694	0.141694	26	46	1.26526664	0.269619676	0.315526	0.269619676	4.669366	0.667261	0.667261	1.105796
N36	MUSTEK	0.060012	0.076174	0.353667	0.416366	0.416366	67.8	61.4	0.042296924	0.021335282	0.003712	0.021335282	1.962487	0.479063	0.503581	0.630193
N37	ADCORP	0.094424	0.056517	0.101634	0.292366	0.292366	102	102	0.352379461	0.151616425	0.16273	0.151616425	2.321069	0.479396	0.479396	0.479396
N38	GOLDREEF	-0.3941	0.276556	0.963132	0.050764	0.050764	16.66667	22.9	0.262967966	0.191656221	0.073451	0.191656221	1.47654	0.322741	0.36546	0.36546
N39	HERCOL	0.069441	0.056262	0.116939	0.099439	0.099439	11.6	13.7	0.102194076	0.057109302	0.17399	0.057109302	1.769447	0.441196	0.461202	0.469605
N40	STEERS	0.119209	0.076069	0.361696	0.403665	0.403665	24.6	24	0.19515777	0.069625194	0.330481	0.069625194	2.17264	0.520613	0.556646	0.977271
N41	BUSBY	0.295383	0.169601	0.100664	0.114901	0.114901	35	33.5	0.134419315	0.102939714	0.146096	0.102939714	1.305606	0.23419	0.236034	0.236034
N42	FRONTRNGE	2.801965	0.127457	0.278418	0.303583	0.303583	27.1	26.6	0.363632114	0.113766725	0.003041	0.113766725	3.197996	0.676391	0.676391	0.676391
N43	BRANDCO	0.035641	0.024193	0.042012	0.06479	0.06478	5.2	5.2	0.006057662	0.003993646	0.067329	0.003993646	1.516753	-0.340697	0.340697	0.463586
N44	SPESCOM	0.241599	0.076621	0.178044	0.186263	0.186263	58.2	60.4	0.056644141	0.033639045	0.078476	0.033639045	1.663661	0.396321	0.442133	0.442133
N45	MNET-SS	0.769653	0.717891	3.167609	7.69625	7.69625	205.6	18.4	0	0	0.107166	0	6.666784	0.710069	0.710069	1.274436
N46	PPC	0.121495	0.097959	0.120644	0.123619	0.123619	373.1	361.6	0.179028479	0.113571641	0.124196	0.113571641	1.576348	0.276661	0.278629	0.326051
N47	ZAPTRONIX	0.378082	0.263636	0.169114	0.423088	0.423088	8.2	8.2	0	0	0	0	2.17106	0.330952	0.330952	0.330952
N48	WETHLYS	0.173149	0.142229	0.168037	0.201048	0.201048	15.6	15.6	0.021400286	0.016127196	0.029283	0.016127196	1.326869	0.248403	0.248403	0.320566
N49	FOSCHNI	0.120839	0.08623	0.118618	0.133026	0.133026	190.1	100.1	0.060190851	0.033390184	0.047906	0.033390184	1.803182	0.211863	0.211986	0.236498
N50	COMPCLEAR	0.411961	0.216161	0.180009	0.187365	0.187365	11.3	11.3	0	0	0.00036	0	1.155744	0.099304	0.099304	0.184978
N51	PALS	0.101437	0.036472	0.316092	0.227299	0.227299	23.2	23.2	0.162935421	0.049565548	0.153327	0.049565548	3.086519	0.651968	0.651968	0.651963
N52	CARGO	0.081878	0.062157	0.176904	0.081181	0.081181	36.3	39.9	0.168015874	0.075575285	0.322086	0.075575285	2.248290	0.496776	0.50664	0.54242
N53	JASCO	0.141546	0.122739	0.324686	0.616133	0.616133	40.1	44.2	0.131011086	0.116621866	0.284478	0.116621866	2.611028	0.632151	0.632151	0.592428
N54	PINNACLE	0.054921	0.016884	0.142004	0.001343	0.081343	3.5	3.7	0.202294901	0.075948286	0.201371	0.075948286	2.863868	0.61144	0.633417	0.633417
N55	LA-GROUP	0.189383	0.141729	0.469548	0.788543	0.788088	68.64532	67.2	0.007660793	0.00324189	0.000805	0.00324189	2.380125	0.578294	0.578294	0.637644
N56	PUTCO	0.047374	0.027938	0.171475	0.186264	0.186264	67.2	55.3	0	0	0.066506	0	2.57744	0.515578	0.515578	0.631193
N57	GROUP-5	0.043424	0.022721	0.269524	0.19108	0.18198	90.4	90.4	0.034660781	0.00942851	0.068655	0.00942851	3.878167	0.710633	0.719833	0.790875
N58	CONCOR	0.003366	0.008002	0.187143	0.036204	0.036204	82.8	79.1	0.072564626	0.024661582	0.12358	0.024661582	2.943832	0.654696	0.663726	0.666055
N59	CONTROL	0.157694	0.090555	0.143473	0.198501	0.198501	14.9	14.9	0.108655141	0.061853695	0.168467	0.061853695	2.114707	0.522461	0.533147	0.660114
N60	DCENTRIX	0.257095	0.070577	0.301665	0.535126	0.535126	11	11	0.783556125	0.189175846	0.003301	0.189175846	4.141946	0.758588	0.758588	0.800943
N61	ENSERVE	0.087082	0.035561	0.090296	0.136949	0.136949	2.6	26.2	0.046766769	0.015002786	0.017665	0.015002769	3.117336	0.396665	0.426825	0.73415
N62	SOFTLINE	0.687636	0.18412	0.157606	0.176942	0.176942	26.8	24.3	0.037272174	0.024422061	0.002004	0.024422061	1.626166	0.22636	0.22636	0.601472
N63	MONEYWB	1.441712	0.225479	0.03751	0.216319	0.216316	-13.4	3.1	0	0	0	0	1.402395	0.255565	0.255565	0.278776
N64	CHEMSERVE	0.111278	0.060639	0.265101	0.291152	0.291152	123.5	118.6	0.97441791	0.030902877	0.153518	0.030902877	2.406093	0.519937	0.519937	0.635289
N65	DATATEC	0.062604	0.03667	0.121759	0.177462	0.177462	20.7	20.7	0.62942767	0.21761661	0.218543	0.21761661	3.613129	0.643555	0.643555	1.936662
N66	WINHOLD	0.033324	0.004627	0.114618	0.040409	0.036204	2.8	2.8	0.636066639	0.142710281	0.201246	0.142710281	4.450197	0.609744	0.609744	0.609744
N67	PASDEC	0.000693	-0.043441	-0.001365	-0.143937	-0.143937	-20.8	-20.8	0.196066269	0.084020055	0.320487	0.084020055	2.357596	0.575741	0.575741	0.635299
N68	OMNIA	0.100036	0.057694	0.400081	0.295374	0.295374	158.2	145.5	0.269467134	0.094157232	0.251062	0.094157232	3.201949	0.664512	0.735963	0.770867
N69	ERP.COM	0.143922	0.109373	0.434656	0.444623	0.444623	4.7	4.7	0	0	0	0	1.66166	0.462909	0.475478	0.549376
N70	MTN	0.166749	0.114134	0.153342	0.217183	0.217163	14.2	14.2	0.579590336	0.269755436	0.256374	0.269755436	2.146577	0.496256	0.607954	1.643664
N71	MATHOMO	-0.057065	-0.067718	-0.225754	-0.416664	-0.416664	-24.5	4.4	0.276160666	0.97730335	0.497664	0.07730335	3.596559	0.722111	0.755963	0.772612
N72	MASONITE	0.02644	0.013614	0.054662	0.030664	0.030664	50	50	0.004475703	0.002873159	0.056496	0.002873159	1.657784	0.291057	0.291057	0.302737
N73	ADONIS	0.093594	0.031631	0.096575	0.055006	0.055006	15.4	15.4	0.163334326	0.095605613	0.125669	0.095029496	1.703074	0.376205	0.376205	0.376205
N74	AFLITEC	0.14196	0.12696	0.143969	0.25957	0.25957	17	17	0.219291911	0.123959476	0.195855	0.123959476	1.760994	0.416601	0.416601	0.432661
N75	GRINTEK	0.054022	0.054663	0.584462	0.714673	0.714673	21.6	20	0.455523955	0.13005345	0.129734	0.13005345	3.50259	0.714412	0.716593	0.731563
N76	DELTA	0.27713	0.146691	0.360428	0.345681	0.345681	212.5	326.2	0	0	0.971243	0	1.966187	0.409486	0.409486	0.742532
N77	AMAPS	0.040639	0.046792	0.348401	0.348401	0.348401	22.9	18.24	0	0	0	0	1.762459	0.436977	0.436977	0.473421
N78	STELLA	0.066646	0.076269	0.04644	0.061194	0.061194	1.2	1.8	0.785826694	0.294110066	0.267127	0.294110066	2.67186	0.625732	0.625732	0.662452
N79	VALUE	0.239528	0.112669	0.213439	0.247675	0.247675	12.7	12.6	0.213437139	0.115767436	0.12215	0.115767436	1.643672	0.36164	0.424341	0.465846
N80	REBSERV	0.156731	0.065416	0.140757	0.243664	0.243664	64.9	64.9	0.709757439	0.272107951	0.004678	0.272107951	2.606367	0.616173	0.620107	0.697124
N81	MT-EAGLE	0.968171	0.056294	0.090126	0.10676	0.10676	42.8	42.9	0.156291766	0.075913759	0.163776	0.075913759	2.056607	0.336435	0.343363	0.344907
N82	AF-&-OVER	0.063184	0.039042	0.06474	0.119667	0.11764	66.4	66.4	0.006523621	0.002849593	0.006623	0.002849593	2.266317	0.161037	0.162651	0.193926
	Max	2.801965	0.717891	5.62661	10.14302	10.14302	599.8	398	1.26526664	0.350982127	0.539786	0.350982127	6.666784	1.078369	1.078369	1.936662
	Average	0.201026	0.103226	0.326211	0.426167	0.42614	66.418	56.18415	-0.942639523	0.062254719	0.123361	0.062229596	-1.12166	0.467211	0.491561	0.636021

All Year 2 Potential Variables Selected [Format: Variable Name (Year Prior to Failure)]

S9(2)	C1(2)	C2(2)	C3(2)	C4(2)	C5(2)	C6(2)	C7(2)	C8(2)	C9(2)	C10(2)	C11(2)	C12(2)	M1(2)	M2(2)	M3(2)	M4(2)	M5(2)	M6(2)	M7(2)	G1(2)	G2(2)	G3(2)	G4(2)	G5(2)	G6(2)	N1(2)	N2(2)	N3(2)	Short Name	Ref Code
0.198329	0.015807	0.042651	0.028207	365.8752	0	0.51508	0.494859	6.331833	-0.48713	-0.00291	8.518557	1.167795	0.103758	0	0.068724	5.236825	1.636068	5.262217	-0.03344	9.322522	6.339832	0.230252	0.916563	0.097457	0.029338	0.186562735	1	1	CAXTON	N1
0.374508	0.055173	0.591785	0.056731	12.99237	0.811596	0.402878	0.320726	2.872024	0.296744	0.013776	0.410994	0	0.240063	0.020064	0.047833	8.233043	3.742663	15.84986	0.045211	9.079179	6.446261	0.540951	0.883313	0.154953	0.564728	0.020362934	0	0	TRUWTHS	N2
0.223895	0.022968	0.093009	0.039565	426	0.011362	0.407046	0.217888	350	2.277426	0.50609	0.204158	0	0.392658	0.033465	0.108299	7.317165	2.11098	11.22358	0.015065	5.471079	5.686367	0.126867	0.895313	0.1319	0	0.377272252	0	0	INVICTA	N3
0.656012	0.034147	0.034147	0.095942	0.583154	2.591602	0.16863	0.068799	-0.66235	-0.48905	0.068734	3.852174	0	0.152213	0	0.06	1.858111	3.755552	3.999543	-1.03611	5.157929	5.273913	0.144416	0.990122	0.746683	0.331344	0.163400641	1	1	AME	N4
0.293956	0.130978	0.149287	0.373002	1.332534	0.149236	0.420317	0.324545	-0.28124	1.091168	0.987094	3.157961	0	0.386136	0	0.098794	13.27682	5.683254	-184.857	-0.27386	5.591248	9.152624	0.426153	1	0.815165	0.011803	0.254019991	0	0	MGX	N5
0.454721	0.049127	0.413848	0.076764	402.8979	0.00243	0.411504	0.289413	1.144144	0.070529	0.002516	0.509534	0.506548	0.11725	0.018908	0.064918	3.317748	2.877702	3.768932	0.047607	5.846293	6.034106	0.318436	0.706288	0.343183	0.496316	0.122997241	0	0	MRPRICE	N6
0.472083	0.024102	0.703719	0.372324	155.2	0	0.387749	0.201388	0.242308	0.179902	0	0.171121	0	0.490119	0.138614	0.038614	0.623219	0.619449	16.96378	-0.2137	5.341698	4.811956	0.177551	-1.161058	-0.02802	0.000564	0.004095616	1	1	HOWDEN	N7
0.50956	0.005745	0.092605	-0.00912	-0.95198	0	-1.44236	-1.59628	0.20048	-1.33772	0.083777	0.801108	0	0.206941	0	-0.58125	2.666306	2.770257	0.580581	-1.36814	4.006295	4.1394	-0.391679	1	2.609727	0	0.290996791	1	1	INFOWAVE	N8
0.402118	0.327874	0.395831	0.30688	1.16697	1.173492	0.196932	0.097763	-0.10169	0.037987	0	0.205531	0	0.156654	0.046421	0.154211	0.237921	0.462439	4.638614	0.127238	6.327694	5.308681	0.129218	1.871268	0.07632	0.049125	0.245001587	0	0	SEARDEL	N9
0.551503	0.019724	0.016724	0.020694	2.780659	0.029092	0.652431	0.55293	-0.23062	0.629792	0.057303	6.783259	0	0.393212	0	0.098171	4.627858	4.953049	4.991573	-0.56884	4.568976	4.615606	0.250661	1	0.465623	2.615645	0.559005716	1	1	EOH	N10
0.391015	0.138227	0.190313	1.880387	0.337089	3.268836	0.023025	-0.03362	-0.4456	0.030922	0	0.658508	0	0.302558	0.018691	0.196747	1.959426	1.447531	8.88184	-0.01297	5.578494	5.462819	0.239904	0.848366	0.123435	0	0.469002074	1	1	NUWORLD	N11
0.441961	0.693336	0.776073	1.560172	9.475113	9.196166	0.099995	-0.05525	0	-0.69494	0	0.129606	0	0.042548	0	0.119372	0.363707	0.436192	5.565609	-0.11671	4.675613	3.869641	0.072676	1	0.071551	0.240365	0	0	0	LABAT	N12
0.443423	0.097073	0.097073	2.011507	0.09561	11.24971	0.046438	-0.04469	-0.35009	0.347031	0.126893	0.38856	0	0.74139	0	0.126571	0.705761	0.662283	0.849046	-0.8099	4.625323	4.120601	0.103907	1	0.757153	0	0.036642105	1	1	AWETHU	N13
0.520664	0.071451	0.470902	0.139725	0.752965	1.516942	0.166714	0.083947	-0.30909	-0.21854	0	0.478475	0.871937	0.418291	0.018173	0.050135	3.568013	3.878752	10.68249	0.199199	6.237338	9.506469	0.163136	0.69441	0.142247	0.481114	0.005996634	1	1	NUCLICKS	N14
0.165177	0.007912	0.457681	0.011163	43.19288	0.257614	0.364794	0.174524	-3.37143	-0.12581	0	0.220973	0	0.068751	0.04902	0.10549	3.175405	0.646866	6.931759	-0.23242	5.292433	5.022322	0.684684	0.536006	0.036582	0.210634	0.575719407	1	1	REX-TRUE	N15
0.686581	0.34965	0.34965	0.491066	1.09717	2.169493	0.185268	0.046772	0.47412	9.032619	0	0.104092	0	0.082632	0.02934	0.967492	1.571188	3.688182	7.821033	0.894846	4.329561	4.367598	0.150155	0.564609	0.155332	0	0.968140351	0	0	SPANJAARD	N16
0.60377	0.22178	0.433551	-0.81429	-0.15558	-1.16484	-0.0359	-0.10501	-0.03958	-0.03495	0	0.22658	0	0.041876	0.109756	0.312195	0.311962	0.475394	6.966006	-0.15757	4.924148	4.199149	-0.138305	0.849505	0.096685	0	0.10940902	1	1	INMINS	N17
0.109847	0	8.32E-05	0	4.02168	0.37354	0.940067	0.963364	0.062355	0.068274	0.01118	11.27179	0	0.069166	0	-0.03612	10.20098	1.28057	2.86585	-0.68554	5.412761	5.46219	0.150209	1	-0.041599	0	0.120062337	0	0	KGMEDIA	N18
0.481241	0.069156	0.111228	0.116199	2.135647	0.24645	0.392477	0.324749	-0.34817	0.250366	0.262537	1.535257	0	0.421816	0	0.05498	3.134015	3.158053	3.680909	-0.37679	5.46763	5.655226	0.102214	1	0.158947	2.740384	0.298084165	0	0	SASANI	N19
0.242948	0.132695	0.339637	0.893201	0.907354	3.961426	0.146031	0.021398	0.036585	-0.04186	0	5.184058	0	0.442616	0.019696	0.13538	3.108621	1.196048	6.148626	-0.29214	6.315725	6.193807	0.578413	0.954494	0.138355	0	0.091503963	1	1	ELLERINE	N20
0.234679	-0.02772	-10.4095	-0.07967	-2.40384	-0.06431	-0.30573	-0.43219	9.530695	-0.72718	0	0.287836	0	0.460394	0	-0.83571	1.308411	0.40188	-5.3646	-1.00527	4.631967	4.119256	-0.120569	1	-0.257061	0	0.491478723	1	1	GLOVIL	N21
0.534993	0.271268	0.493592	0.133946	4.661564	0.752625	0.499662	0.32953	-0.1347	0.476028	0	0.171735	0	0.075236	0.874627	0.160597	0.730472	0.948361	5.615298	0.42148	5.535368	5.150937	0.268864	0.595171	0.104379	0.736072	0.193920921	1	1	BEARMAN	N22
0.233623	0.001295	0.062915	0.091784	2.99927	0.091966	1.054331	1.265329	0.222997	0.783045	0	2.063716	0	0.143938	0	0.043108	16.46325	5.643411	17.06043	0.662629	5.176184	5.61185	0.262733	1	0.313455	0	0.49162298	0	0	UCS	N23
1.078389	0.060361	0.368304	0.071369	426	0	0.271002	0.240567	350	0.05128	0	0.119183	0	0.727448	0	1.107692	0.650198	-8.922	4.230422	-0.50998	5.115154	4.980974	-0.206702	1	-2.666959	1.020786	0.090639063	1	1	CONNECT	N24
0.505076	0.007918	0.02642	0.008996	69.70358	0.040828	1.049135	1.019362	-10.1408	1.040428	0.060694	11.00213	0	0.363246	0.095923	0.111994	7.405629	8.462316	5.199503	0.665227	6.682389	7.255308	0.192972	0.944649	0.847099	0.066667	0.079760715	0	0	COMPAREX	N25
0.389625	0.223635	0.235357	0.193658	0.623444	0.603074	0.510181	0.336248	-0.00053	-0.04834	0	0.214659	0	0.100312	0.026186	0.068676	3.652724	2.479269	10.52661	-0.43656	5.584605	5.741019	0.211016	0.597187	0.100724	0.173968	0.195958111	0	0	CITYLDG	N26
0.570746	1.163181	1.064711	0.515635	0.582253	3.963399	0.141507	-0.00255	-0.26193	-0.21306	-0.02986	0.201663	0	0.114621	0	-0.00505	0.352461	0.479512	9.072334	-0.40981	5.170763	4.474317	0.380965	1	-0.091818	0.005533	0.394845224	1	1	TRNPACO	N27
0.399602	0.350462	0.350462	0.40678	1.038578	1.233166	0.290261	0.111411	0.711934	-0.69416	0	0.179147	0	0.16313	0.222222	0.372222	0.491634	0.342371	4.239392	-0.34716	5.264412	4.549757	0.360549	0.42424	0.054518	0	0.031751927	0	0	ARGENT	N28
0.593433	0.443126	0.443705	0.303829	1.426621	1.968756	0.236733	0.143739	-0.28621	0.216571	0	0.466693	0	0.063193	0.029586	0.111934	0.981953	1.015711	6.791509	2.126539	4.473545	4.060672	0.255663	0.682292	0.077351	0	0.690947368	0	0	NICTUS	N29
0.48404	0.005138	0.040406	0.00706	5.594568	0.061967	0.593936	0.60214	-0.4449	9.654499	0.079777	1.491763	0	0.250043	0	0.87449	4.916312	4.623063	4.663047	0.33332	5.829275	6.205795	0.188571	1	0.302751	0.251354	0.207109007	1	1	AST GROUP	N30
0.701240	-1.37139	-1.92076	0.275183	0.653693	1.601061	0.37677	0.276026	-0.11256	0.358126	0	0.057097	0	0.149748	0	-0.44516	0.145032	-30.4807	-101.669	-0.52235	5.379454	4.439231	-0.511349	1	10.14302	0	0.156761267	1	1	SEKUNJALO	N31
0.442234	0.211153	0.294533	0.257987	0.905041	2.109602	0.243131	0.224692	-0.18542	-0.089	0.022373	1.092767	0	0.223816	0	0.042939	3.489903	2.997637	10.78367	0.114079	6.878874	7.06854	0.157917	1	0.127203	0.031187	0.068262201	0	0	IMPERIAL	N32
0.300965	0.656758	9.658625	0.593491	0.262746	2.514716	0.19945	-0.0991	0.031436	-0.02538	0	0.057143	0	0.06375	0.044231	0.069231	1.609366	0.695295	25.40919	-0.71881	5.964037	5.669212	0.011605	0.114586	0.005467	0	0.328330332	0	0	MOBILE	N33
0.513268	0.055576	0.187126	1.531955	6.920613	1.695489	0.018349	-0.07233	-4.53846	-0.37532	0	1.60973	0	0.388893	0	0.110643	2.465066	2.669148	5.58796	-0.79677	4.751964	4.654137	0.25936	1	0.298373	0	0.479450941	1	1	MAXTEC	N34
0.667261	0.407061	0.575173	0.366996	0.591292	1.977729	0.239096	0.145496	0.252623	-0.03958	0.051461	9.76393	0	1.009931	0	0.05501	1.003885	3.1412	8.035339	-0.55563	9.191648	6.017628	-0.026645	1	0.141584	0.734364	0.154264904	1	1	PRIME	N35
0.479063	0.078467	0.148794	0.05712	13.90772	0.01251	0.619368	0.369713	-0.72432	0.402625	-0.04369	4.130609	0	0.42055	0	0.060064	5.003599	4.7521	12.77092	-0.29298	5.610521	5.990196	0.360226	1	0.418396	0	0.271607632	0	0	MUSTEK	N36
0.479396	0.107914	0.125043	0.066966	1.83312	0.584736	0.560618	0.437871	0.256749	0.873717	0.653231	1.621167	0	-0.357007	0.008538	0.054429	5.037779	5.605641	12.63909	0.040665	6.401506	5.784439	0.192731	0.829784	0.242618	0	0.137729344	0	0	ADCORP	N37
0.322741	0.614314	0.614314	-0.2062	-0.36598	-1.84716	-0.21733	-0.03848	1.573535	1.026046	0.40143	361.8376	0	0.064397	0	0.165017	0.854648	0.407273	-12.2408	-0.85829	8.663619	5.12426	0.099951	1	0.050784	0	0.652825722	1	1	GOLDREEF	N38
0.441168	0.196569	0.217116	-0.19558	-1.58377	-1.92365	-0.20502	-0.32154	-0.72717	-0.67613	0	0.307014	0	0.205369	0	0.063453	1.462264	1.15436	19.58659	-0.18786	6.26235	5.99196	0.27945	1	0.099439	0	0.639627697	1	1	HERCOL	N39
0.520513	0.149424	0.16969	0.213156	2.543792	1.446239	0.436963	0.183066	1.066106	0.191202	0.002432	0.060451	0	0.207996	0.074634	0.162795	1.617304	2.055166	3.775361	0.658634	5.027423	5.003262	0.248127	0.521358	0.210449	0	0.403000615	1	1	STEERS	N40
0.23419	0	0.00733	0	0.637975	2.224624	0.280428	0.201627	0.973484	1.099786	0.243436	1.644511	0	0.307653	0	0.136187	2.610909	0.859593	3.817057	-0.61752	5.464626	5.263242	0.119624	0.963184	0.099179	0	0.335002276	0	0	BUSBY	N41
0.676391	0.01443	0.020525	0.122745	2.192693	0.012168	0.36881	0.340726	0.352812	0.351057	1.975774	28.81169	0	0.652135	0	0.016064	8.551255	26.66028	1.746161	0.22735	5.576617	6.386675	0.111966	1	0.303583	0	0.266162523	0	0	FRONTRNGE	N42
0.340697	0.149051	0.364176	8.209184	0.592742	36.88946	0.006946	-0.09363	2.704637	-0.19827	0.027397	1.094181	0	0.417209	0	0.12093	0.966266	0.499321	7.894076	-0.66817	5.39512	4.912585	0.183661	1	0.06476	0	0.165251165	0	0	BRANDCO	N43
0.398321	0.061196	0.09287	0.62355	1.11496	2.039018	0.694161	-0.00618	0.921169	-0.25159	0.186003	1.414097	0	0.417487	0	0.057241	3.490681	2.341419	3.991187	-0.35744	5.534496	5.677663	0.149642	1	0.168263	0	0.069217919	1	1	SPESCOM	N44
0.710069	0.016096	0.016645	0.063076	420	0.430726	0.350363	-1.30467	350	0.194634	0.017992	0.748656	0.960149	0.043037	0.57037	0.423457	3.006125	16.96355	2.850605	-0.09963	5.809232	6.138407	-0.137628	-0.346941	-2.77423	2.619911	0.016000565	0	0	MNET-SS	N45
0.275631	0.256106	0.27907	0.138672	1.564207	0.690291	0.65033	0.506637	-0.11415	0.100937	0	0.205154	0	0.17096	0.059165	0.061764	3.504379	1.526306	5.119356	-0.01542	9.371567	8.358125	0.34466	0.254144	0.031468	0.066625	0	0	0	PPC	N46
0.330952	0.022229	0.100043	0.023406	420	0	0.797327	0.75771	350	0.062787	0.156709	1.897542	0	0.369667	0	0.221429	2.129624	1.530327	2.471264	-0.71692	4.582186	4.430297	0.371621	1	0.423066	0	0.612443662	1	1	ZAPTRONIX	N47
0.246403	0.003765	0.089697	0.008016	6.442561	0.281643	0.421669	0.161438	-0.11298	0.145615	0	0.164603	0	0.146946	0.021429	0.060612	7.627947	2.494102	10.62312	0.006736	5.119596	5.393649	0.162669	0.734189	0.147606	0	0.722210345	0	0	WETHLYS	N48
0.211663	0.073171	0.440332	0.066667	3.064699	0.466148	0.483005	9.207692	0.707151	0.415564	0	0.404435	0.741097	0.301694	0	0.069179	6.063095	1.937247	6.066827	-0.1176	6.421357	6.531536	6.479694	0.842949	0.085529	0.202998	0.030638298	1	1	FOSCHINI	N49
0.099304	0	0.039679	0	426	0.001609	2.254534	1.044377	350	-0.36094	-0.05503	0.431091	0	0.153732	0.018644	0.095763	16.52321	1.89636	3.35553	-0.77248	4.522601	4.73766	0.259786	0.911198	0.170745	0.171696	0.413203515	1	1	COMPCLEAR	N50
0.651963	0.32635	0.32635	0.871994	2.076775	1.469526	0.157887	0.031762	-0.05822	0.170485	0	0.397615	0	0.2924	0.045	0.232	0.486405	0.878474	3.540199	1.060243	4.498779	4	0.182311	0.848892	0.192953	0	0.5797	0	0	PALS	N51
0.49217	0.578848	0.578848	0.536317	1.639706	2.59895	0.249955	0.132867	0.095654	-0.34864	0	0.326799	0	0.0134	0.066667	0.263704	0.256847	0.266071	3.409579	0.006865	5.326743	4.431364	0.410191	0.765074	0.062109	0.045592	0.67615	0	0	CARGO	N52
0.632151	0.221666	0.595905	0.079086	6.025654	0.397328	1.132606	1.012028	-0.74926	-0.06769	0.064487	6.367515	0	0.712992	0.042403	0.173496	1.591612	2.826267	5.799643	0.283041	5.064952	5.06761	0.283114	0.711463	0.36721	0.576559	0.309799457	0	0	JASCO	N53
0.61144	0.343015	0.411619	0.777096	0.600224	4.417512	0.074553	-0.02324	1.967065	0.225411	0.058618	-0.96414	0	0.417043	0	0.291667	0.173106	0.261934	5.494147	-0.13668	5.215779	4.240447	0.027952	1	0.061343	0	0.23349813	1	1	PINNACLE	N54
0.576294	0.009493	0.109954	0.987563	159.9703	0.091168	0.898156	0.859389	1.524324	0.485387	0	0.833781	0	0.368931	0.020279	0.064468	6.742551	8.170712	19.77617	0.244479	5.05743	5.846898	0.33237	0.747252	0.587749	5.725528	0.330077743	0	0	LA-GROUP	N55
0.515578	0.080524	0.080524	0.665238	426	0.522518	0.321109	0.201741	350	0.218508	0	0.096242	0	0.090142	0.320442	0.371271	0.377524	0.561681	3.460183	-0.21742	5.390455	4.879694	0.245369	0.165014	0.030736	0.200667	0.111447498	0	0	PUTCO	N56
0.710533	0.300264	0.338219	0.558979	8.379534	1.443992	0.084654	0.046673	-0.02495	-0.05079	0	0.453609	0	0.261387	0.066957	0.327536	0.186338	0.491946	5.787952	-0.57862	9.191106	5.307628	0.062931	0.710021	0.114356	0.326637	0.006102792	0	0	GROUP-5	N57
0.654698	0.765884	0.798912	0.184343	6.013798	0.833255	0.226531	0.162676	1.92118	0.28414	0	0.409184	0	0.365148	0	0.042764	0.439549	0.847094	16.39361	-0.4237	5.756273	5.215321	0.23519	0.319412	0.011584	0.016672	0.006143242	0	0	CONCOR	N58
0.522461	0.106153	0.152659	0.431	1.027365	3.162361	0.101967	0.061363	0.925605	-0.35636	0	1.204556	0	0.309586	0.1	0.16625	0.966149	1.967431	3.115033	-0.38623	5.146875	4.849665	0.191315	0.241265	0.047891	0.625496	0.374895464	1	1	CONTROL	N59
0.766851	0.091563	0.02025	0.002698	1.436146	0.012133	0.356653	0.342101	-0.99973	-0.00313	-0.2514	0.935364	0	0.23005	0	0.037415	4.546891	14.29237	5.22168	-0.04962	5.132628	6.870528	0.221471	1	0.535128	0	0.20442974	1	1	DCENTRIX	N60
0.396665	0.102048	0.136952	0.046728	16.86343	0.069903	0.637814	0.535959	0.667605	0.062179	-0.04593	0.401584	0	0.440629	0.066119	0.021468	1.062336	1.313622	3.395569	0.736926	6.491553	5.106239	0.128169	0.306675	0.042612	0.525792	0.263470047	1	1	ENSERVE	N61
0.22838	0.017265	0.017259	0.119212	3.355698	0.024458	0.356846	0.356239	0.186632	0.71264	0.421784	17.64207	0	0.446604	0	0.029712	18.26154	6.364991	4.872237	1.732964	6.898129	6.516325	0.165761	1	0.176942	11.48075	0.311236282	0	0	SOFTLINE	N62
0.255565	0.000163	0.017618	0.091049	420	0	0.707993	0.686146	350	0.356347	0	1.024706	0	0.192125	0	-0.46207	5.741153	2.05765	0.812672	-0.61171	4.196967	4.365486	3.591513	1	0.216316	0.025395	0.483325	1	1	MONEYWB	N63
0.519937	0.176322	0.290672	0.146827	6.766016	0.565266	0.522323	0.293618	-0.38249	0.228431	0.090427	0.096572	0	0.134771	0.026209	0.10475	2.257376	2.826365	6.445696	-0.20019	5.646461	8.916916	0.336514	0.73913	0.215199	0.21346	0.091430533	1	1	CHEMSERVE	N64
0.643555	0.238118	0.276901	1.057572	0.217061	4.626703	0.073366	-0.04636	-0.5724	-0.16047	0.242141	3.710725	0	0.236579	0	0.022403	6.200631	15.21609	36.26505	3.369327	4.70559	5.306612	0.072644	1	9.177462	10.57244	0.474646066	0	0	DATATEC	N65
0.609744	0.699147	0.763115	1.197375	0.315265	4.472669	0.073792	-0.02197	0.024626	-0.00687	0	0.151393	0	0.049149	0	0.164706	0.96791	0.249146	6.519281	-0.83168	5.320431	4.049616	-0.149692	1	0.040409	0	0.195987444	1	1	WINHOLD	N66
0.575741	6.60464	3.761744	0.780532	0.920222	4.145096	0.134261	0.042695	0.647635	0.092274	0.009204	0.154759	0	0.161568	0	-0.26889	0.378932	0.814349	11.41623	0.417569	4.965426	4.304211	-0.060479	1	-0.143937	0.145749	0.611035666	0	0	PASDEC	N67
0.684512	0.405341	0.437259	0.586291	1.400179	2.130112	0.172145	0.032787	-0.42179	-0.01059	0	0.340252	0	0.066666	0.042618	0.109254	1.233442	2.703423	7.940605	9.317002	5.822709	5.749208	0.182643	0.602501	0.177963	0.115121	0.198013672	0	0	OMNIA	N68
0.462909	0	0.104175	0	420	0	0.54721	0.452659	350	0.582385	0	0.700074	0	0.071106	0	0.261111	1.961977	1.706227	3.834531	-0.64896	4.500679	4.463173	0.27292	1	0.444523	0	0.127999554	1	1	ERP.COM	N69
0.496258	0.176213	0.234358	0.152016	0.806106	1.176078	0.437506	0.397064	-0.36767	-0.08424	0	0.399361	0	0.066775	0.003623	0.01715	6.61916	9.227232	14.9269	0.52189	5.978867	6.611967	-0.038022	0.788786	0.171296	1.161055	0.005301006	0	0	MTN	N70
0.722111	-0.47301	-0.62469	0.215561	2.13968	2.94965	0.229079	0.152716	0.651312	0.158324	0	0.199349	0	0.193461	0	-3.0625	0.052443	0.136277	-12.929	-0.95493	5.18563	3.763925	-0.413908	1	-0.416884	0	0.187098152	1	1	MATHOMO	N71
0.296483	0.426991	0.50344	0.154132	43.44296	0.452624	0.426845	0.333657	15.47347	0.281115	0	0.108429	0	0.049868	0.025243	0.097887	0.707375	0.320723	4.865893	-0.43856	5.231836	4.545464	0.605792	0.737792	0.022771	0.024167	0.000586682	1	1	MASONITE	N72
0.375329	0.374266	0.646179	0.406769	0.953715	1.376117	0.24313	0.046284	-0.32704	0.25555	0.005841	0.015781	0	0.068136	0.012712	0.085254	1.32921	0.85099	5.921429	0.636413	4.233529	3.93222	0.453011	0.802893	0.044166	0	0.700413793	0	0	ADONIS	N73
0.416801	0.092665	0.233427	0.097683	1.931694	0.517646	0.574569	0.63427	1.405815	0.273452	0.393342	2.436738	0	0.633775	0.005034	0.987047	6.038283	4.432	10.66654	0.003439	5.325171	5.728013	0.151631	0.924635	0.240008	0	0.193218158	0	0	APLITEC	N74
0.714412	0.122095	0.294552	0.267656	1.115665	0.694106	0.203102	-0.00481	-0.05592	0.732069	0.01422	-0.70662	0.973223	0.353621	0.031452	0.174194	0.966663	2.46942	3.071914	-0.61315	5.563202	6.411408	0.045021	-3.451926	-2.467667	0.117132	0.008684237	0	0	GRINTEK	N75
0.409466	0.057236	0.079694	0.06438	420	0.213489	0.614845	0.47036	350	0.0717	0	0.537156	0	0.138063	0.034292	0.056133	5.226641	3.991766	6.607606	0.501252	5.939957	9.270419	0.366516	0.612812	0.21196	0.862617	0.009518773	1	1	DELTA	N76
0.436977	0	0.136978	0	420	0	-0.17684	-0.15295	350	0.189104	0.504545	13.79671	0	0.670695	0.910345	0.096279	4.541334	3.553406	12.33407	0.933333	5.043586	5.343212	0.056478	0.694491	0.311642	0.030319	0.463157695	0	0	AMAPS	N77
0.925732	0.040344	0.173423	-0.00549	-1.23115	-0.73771	-0.57669	-0.58504	-0.99731	-0.26135	0	0.517033	0	0.156366	0	0.031579	1.56469	2.616308	30.56436	-0.22644	4.73504	4.725912	0.0425	1	0.061194	0	0.655135714	1	1	STELLA	N78
0.38184	0.077633	0.077633	0.876858	2.500066	0.422042	0.758375	0.588999	-0.52661	-0.09522	0	0.497196	0	0.174156	0	0.154676	2.272112	1.5987	2.689971	-0.81436	5.292956	5.231039	0.331313	1	0.247875	0.061627	0.539997784	1	1	VALUE	N79
0.618173	0.006673	0.075379	0.011992	0.517095	0.033248	0.226354	0.240216	-0.44277	0.481178	0.275564	6.963759	0	0.255492	0.095561	0.027465	6.874917	9.763632	12.12082	1.394547	9.012595	9.595836	0.133225	0.920543	0.224487	0.096307	0.187315877	0	0	REBSERV	N80
0.338435	0.348633	0.348633	0.870649	0.536927	4.506667	6.120437	-0.0433	0.06992	0.62511	-0.00942	0.456481	0	0.006625	0.012143	0.060657	2.563357	1.766076	14.31333	-0.93631	4.708184	4.644439	0.09053	0.776574	0.062907	0.003968	0.603609524	1	1	MT-EAGLE	N81
0.181037	0.967339	0.317216	0.007275	44.96364	0.068833	0.706059	0.576241	0	0.284192	0	0.417838	0	0.196716	0.050347	0.153472	1.677065	0.777961	4.255266	-0.00238	5.26558	4.916832	0.381032	0.615695	0.097612	0.063947	1	0	0	AF-&-OVER	N82
1.078389	6.60464	3.761744	5.209184	420	36.88946	2.254534	1.844377	350	2.277426	1.975774	361.8376	1.167795	1.009931	0.57037	1.107692	18.46325	20.66028	36.25505	3.369327	9.978074	7.255308	3.591513	1.871266	10.14302	11.48075	1	1	1	Max	
0.467156	0.243043	0.156118	0.335616	68.7355	1.703575	0.33679	0.218552	42.97694	0.150016	0.09166	6.56116	0.061238	0.267196	0.032481	0.049092	3.477556	2.663147	3.549259	-0.06109	5.366424	5.264677	0.218232	0.730417	0.223796	0.573238	0.285035169	0.487805	0.487805	Average	

Appendix F.6.

All Year 3 and Other Potential Variables Selected [Format: Variable Name (Year Prior to Failure)]

Ref Code	Short Name	L1(3)	L2(3)	L3(3)	L4(3)	L5(3)	L6(3)	L7(3)	L8(3)	L9(3)	L10(3)	L11(3)	E1(3)	E2(3)	E3(3)	P1(3)	P2(3)	P3(3)	P4(3)	P5(3)	P6(3)	P7(3)	S1(3)	S2(3)	S3(3)	S4(3)	S5(3)	S6(3)	S7(3)	S8(3)	S9(3)	C1(3)	C2(3)	C3(3)
N1	CAXTON	2.508842	2.21938	2.178982	2.512398	2.22192	0.357468	3.91318	11.49806	4.086797	43.45153	1	0.851652	3.788162	1.637024	0.13127	0.123867	0.171123	0.202921	0.202867	470	490	0.003724	0.002185	0.034712	0.002165	1.720192	0.287092	0.267092	0.307968	0.286802	0.027828	0.072133	0.068066
N2	TRUWTHS	2.319193	1.770769	1.6028	2.317081	1.78704	0.148902	4.436774	6.435007	23.89828	105.6796	0	1.704588	0.952072	2.834146	0.098348	0.082281	0.231291	0.233188	0.233198	33.6	34.1	0.051488	0.032056	0.115513	0.032056	1.808141	0.373084	0.396863	2.136447	0.373225	0.03347	0.501659	0.044161
N3	INVICTA	1.575261	0.535824	0.535824	1.576281	0.535824	0.020214	0.338738	4.480824	5.438084	53.4318	0	2.679961	31.77049	7.377381	0.090806	0.084734	0.428134	0.825111	0.825111	53	83	0.173623	0.063035	0.029247	0.063035	2.752708	0.638858	0.638858	0.891803	0.638858	0.091826	0.183813	0.198332
N4	AMS																																	
N5	MGX	1.542049	1.128746	1.128746	1.542049	1.128746	0.855040	2.721162	9.992292	3.193839	80.06364	0	0.887773	2.947344	1.084808	0.846361	0.126839	0.142708	0.141175	0.141176	88.8	88.1	0.138048	0.088718	0.108978	0.088716	1.581847	0.358883	0.358893	0.445467	0.358883	0.024881	0.027858	0.218984
N6	MRPRICE	1.521027	0.600834	0.591488	1.520703	0.60031	0.050588	10.47708	0.185	4.870232	30.28828	1	2.088834	14.43878	3.722062	0.053744	0.034808	0.141277	0.128747	0.129708	25.1	25.1	0.601008	0.000565	0.00048	0.000451	1.780847	0.438501	0.501872	0.585422	0.438448	0.068351	0.025424	0.056558
N7	HOWDEN																																	
N8	INFOWAVE	2.865401	2.865401	2.865401	2.865401	2.865401	0.804822	4.805003	50	4.527284	1.528845	0	0.78428	9.738808	1.77481	0.412852	0.380203	0.135175	0.655443	0.655443	8	8	0.898458	0.290077	0.230818	0.290077	2.263038	0.558119	0.558118	0.558119	0.558118	0.000898	0.008147	0.001254
N9	SEARDEL	2.055335	1.078533	1.078533	2.055336	1.078533	0.115004	0.197310	4.622595	4.887203	81.73073	0	1.405411	4.09898	5.75808	0.86771	0.021257	0.07848	0.122416	0.122416	47.6	45.8	0.281818	0.871018	0.084818	0.871018	4.097775	0.388502	0.457894	0.508548	0.388502	0.224704	0.275783	0.188890
N10	BOH	3.38581	3.384983	3.384983	3.38581	3.384983	0.588148	4.331258	6540.2	21.82755	67.45036	0	1.332868	53.43301	2.862315	0.818901	0.204453	0.30808	0.005871	0.005871	12.2	12.2	0.883828	0.298839	0.000734	0.298839	2.222846	0.947928	0.547926	0.723863	0.647926	0.000763	0.000763	0.005407
N11	NUWORLD	3.018387	2.130073	2.130073	3.018887	2.130073	0.288463	4.8444	0.568718	5.986543	72.86251	0	1.770858	21.38108	2.885507	0.058788	0.042748	0.13818	0.122498	0.122498	105.8	165.8	0.067838	0.094188	0.061487	0.054186	1.61842	0.358255	0.358256	0.418051	0.358255	0.193293	0.247998	0.271662
N12	LABAT																																	
N13	AWETHU																																	
N14	NUCLICKS	1.458774	0.47724	0.47724	1.458774	0.47724	0.17181	47.18069	5.098521	5.341238	11.01583	0	2.315476	18.8876	5.172388	0.052495	0.037536	0.100816	0.104147	0.104147	44.8	44.8	0.17595	0.078788	0.064579	0.078788	2.233835	0.541167	0.546938	0.638238	0.541167	0.11789	0.542507	0.07477
N15	REX-TRUE																																	
N16	SPANJAARS	0.673073	0.015612	0.580272	0.869485	0.812004	0	5.732231	12.81843	5.831177	27.10835	0	2.103443	0.620828	7.898263	0.075761	0.038244	0.380038	0.30203	0.30203	24.9	28.7	0.345088	0.084317	0.295398	0.074251	3.858888	0.726893	0.726893	0.726893	0.708242	0.425175	0.429175	0.801001
N17	INMINS	1.90642	1.18689	1.18698	1.90642	1.18698	0.09179	0.507333	0.874782	7.150261	48.17121	0	2.647474	25.83584	7.101251	0.038122	0.033282	0.215078	0.236341	0.236341	18.4	16.4	0.381247	0.152873	0.618211	0.152873	2.493878	0.598845	0.598845	0.603474	0.598845	0.188265	0.35821	0.181835
N18	KGMEDIA	2.183878	2.183879	2.143474	2.193878	2.183878	0.072558	4.284284	50	9.742888	37.24881	0	0.284512	0.381728	0.312858	0.15088	0.196083	0.050035	0.061388	0.061388	19.78	19.725	0.092871	0.002428	0.615888	0.002436	1.183156	0.065081	0.066081	0.220805	0.066081	0.065738	0.167289	0.067283
N19	SASANI	0.81656	0.780214	0.780214	0.81658	0.780214	0.0197	2.051832	14.8171	4.075802	83.06805	0	0.468844	1.532688	0.747088	0.540224	0.430678	0.274818	0.321302	0.321302	29.78	3.78	0.206777	0.128414	0.151488	0.129414	1.590068	0.361961	0.361961	0.47819	0.381961	0.113378	0.231445	0.32488
N20	ELLERINE	2.627214	2.530531	2.415515	2.827214	2.530531	0.638137	1.734829	13.11231	17.88847	217.4887	0	0.878888	7.808873	1.45057	0.191349	0.11718	0.187448	0.188877	0.188877	246.4	244.6	0.082333	0.055844	0.153788	0.055844	1.850445	0.281888	0.282832	0.610383	0.291888	0.110814	0.334562	0.465629
N21	GLOVIL	2.534888	2.493118	2.493118	2.534888	2.493118	0.287855	7.221451	84.87458	20.14436	38.73528	0	0.474884	1.384749	0.801788	0.471878	0.403878	0.203735	0.24311	0.24311	8.6	9.9	0.083044	0.085458	0.004708	0.085458	1.269647	0.199244	0.199644	1.002811	0.199244	0	0.226939	0
N22	SEARMAN	1.435357	0.485214	0.485214	1.435357	0.405214	0	5.433754	2.774878	0.225588	140.0808	0	1.481701	7.717801	3.833053	0.128839	0.038538	0.28127	0.140002	0.140002	5.3	23.8	0.107178	0.043122	0.34157	0.043122	2.485497	0.587525	0.619172	0.888881	0.587525	0.372186	0.510884	0.505048
N23	UCS	5.794989	5.742313	5.742313	5.794989	5.742313	0.582788	3.298988	65.08328	9.857258	55.85818	0	0.424733	0.731717	0.581045	0.488808	0.34288	0.191291	0.182427	0.182427	9.1	9.1	0.198825	0.118723	0.817918	0.118723	1.320834	0.242842	0.242842	0.28583	0.242842	0.058708	0.164238	0.095599
N24	CONNECT																																	
N25	COMPAREX	1.730485	1.321705	1.180748	1.730485	1.321708	0.178034	3.563702	0.536944	3.025788	37.82745	0	1.087494	35.08188	1.804864	0.11257	0.07849	0.168858	0.145711	0.145711	70.1	64.8	0.016865	0.016312	0.025738	0.010312	1.751701	0.417262	0.417262	0.521406	0.417262	0.143837	0.238845	1.644921
N26	CITYLDG	1.111187	1.068733	1.068733	1.111197	1.068733	0.03482	0.107086	242.0578	7.025823	-18.3639	0	0.378106	0.558388	0.715848	0.508843	0.308564	0.174008	0.220823	0.220823	128.7	128.7	0.770408	0.403808	0.144958	0.403808	1.907881	0.472885	0.472885	0.676513	0.472885	0.229529	0.244922	0.229994
N27	TRNPACO	1.807423	1.05028	1.042297	1.807423	1.05028	0.000281	4.29838	9.037588	0.88297	75.80117	0	1.679557	4.288887	3.348289	0.117815	0.131005	0.408805	0.441322	0.441322	64	59	0.220688	0.110858	0.212034	0.110858	1.883554	0.405704	0.485704	0.531491	0.405704	0.182771	0.327108	0.219352
N28	ARGENT	2.051306	1.338242	1.338242	2.051306	1.338242	0.134045	4.838242	5.723645	4.837783	88.54404	0	1.183885	3.01918	2.100958	0.116442	0.08345	0.20337	0.196334	0.196334	6.92	8.8	0.18142	0.108758	0.113424	0.108758	1.780081	0.401233	0.401233	0.401233	0.401233	0.193587	0.193567	0.239411
N29	NICTUS	1.477553	0.823122	0.517486	1.477583	0.823122	0.000847	5.806782	7.472211	10.721	75.6682	0	2.286328	7.777877	5.178612	0.041488	0.037208	0.259958	0.192603	0.162803	14.3	14.3	0.192087	0.804828	0.329572	0.804828	2.284158	0.552377	0.552377	0.552377	0.562377	0.581115	0.590317	0.439723
N30	ABT GROUP	1.555857	1.536328	1.458475	1.555857	1.536325	0.287088	0.511555	286.8527	7.411195	0.076807	0	2.884027	13.97908	4.919308	12.34481	0.15758	0.778546	0.774612	0.774612	20.8	20.8	0.001941	0.001052	0.00204	0.001052	1.845441	0.457845	0.457845	0.851132	0.487845	9.81E-05	0.000758	0.007741
N31	SEKUNJALO	0.884148	0.549934	0.549834	0.884148	0.549934	0.033312	4.40234	0.263878	0.087318	91.31778	0	1.199875	4.027568	-7.808941	-0.408212	-0.385021	-0.588848	2.777428	2.777428	-136	12.1	-1.80346	0.28287	0.425448	0.28287	-0.37331	0.838878	0.838878	0.838878	0.838878	-0.13417	-0.16877	0.271715
N32	IMPERIAL	1.657726	1.110780	0.820858	1.657728	1.110789	0.061085	18.22881	7.958264	0.143131	22.13183	0	1.380831	5.758449	2.168112	0.060786	0.051858	0.112487	0.112381	0.112381	217.9	218.1	0.005273	0.003317	0.870833	0.003317	1.580438	0.315872	0.325782	0.359583	0.315872	0.078586	0.234841	0.069048
N33	MOBILE	0.828204	0.828204	0.88800	0.828204	0.828204	0.061588	1.178805	50	6.250214	275.9197	0	0.079875	1390.352	0.116788	0.307128	2.231033	0.115037	0.280814	0.280614	16.9	17.8	0.308808	0.211562	0.14516	0.211502	1.486585	0.318245	0.315245	0.315245	0.315245	0.954283	0.95434	0.517394
N34	MAXTEC	1.387279	0.843197	0.843187	1.387278	0.843197	0.2165	4.598271	5.87208	4.097087	52.44229	0	1.501488	82.47584	3.584252	0.744005	0.123858	0.439887	0.441811	0.441611	6.3	6.1	0.007111	0.002988	0.005108	0.002988	2.373708	0.578734	0.578734	0.87188	0.578734	0.004224	0.014839	0.087584
N35	PRIME	0.891836	0.848188	0.845188	0.891836	0.845188	0.14487	4.317888	67.27287	12.32127	61.28882	0	1.182887	4.234124	18.84315	1.867221	0.983535	2.035178	17.53211	17.53211	88	88	0.858808	0.038138	0.158383	0.039138	18.77831	0.481612	0.871084	1.24718	0.481612	0.043851	0.088388	0.331005
N36	MUSTEK																																	
N37	ADCORP																																	
N38	GOLDREEF	0.936876	0.584152	0.589348	0.936876	0.584152	0.028418	4.851164	0.57863	7.581944	81.05350	0	0.477380	31.84987	0.705815	0.279406	-0.116985	-0.030482	-0.08257	-0.08257	-16.0286	-3.42857	0.010888	0.007237	0.154804	0.007237	1.478483	0.219062	0.219062	2.328188	0.2361	0.238576	0.236578	5.890957
N39	HERCOL	1.761113	1.585749	1.458026	1.761113	1.585749	0.035953	1.614404	12.78225	0.868737	202.4054	0	1.187382	13.4586	2.488073	0.407875	0.385738	0.95987	0.679711	0.678711	28.44	29.38	0.023585	0.01188	0.067282	0.01188	2.132481	0.52738	0.533887	0.555183	0.52738	0.013223	0.017083	-0.87359
N40	STEERS																																	
N41	BUSBY	2.861852	1.784747	1.784747	2.861852	1.784747	0.218892	4.274383	2.420358	3.825581	148.2171	0	0.552525	13.1885	0.75843	0.208564	0.155765	0.087721	0.118293	0.118293	25.9	25.9	0.107906	0.078508	0.031782	0.078508	1.374485	0.272444	0.272444	0.272444	0.272444	0.014388	0.021614	-0.0132
N42	FRONTRNGE	6.297189	6.207497	5.148958	6.297188	6.207497	0.540779	2.811788	68.35158	14.07318	118.318	0	0.488587	0.85188	4.282888	0.205824	0.238019	0.054238	1.018811	1.010811	10.8	10.9	7.420146	0.911812	0	0.911812	6.140227	0.880547	0.880547	4.958928	0.880547	0.071472	0.087445	-0.89588
N43	BRANDCO	2.740543	1.713111	1.713111	2.740543	1.713111	0.162537	2.74418	3.178208	4.158783	188.1894	0	1.138847	25.15838	1.781147	0.102911	0.08585	0.00044	0.188404	0.188454	18.6	18.9	0.008338	0.008028	0.058723	0.008028	1.549015	0.354428	0.354428	0.428197	0.354428	0.068081	0.175288	-0.1981
N44	SPESCOM	1.421825	1.188873	1.188873	1.421825	1.188873	0.241842	3.087828	11.21134	3.281708	22.85202	0	1.24208	11.21643	2.780388	0.100487	0.080031	0.171338	0.222519	0.222519	11.8	64.2	0.085481	0.042858	0.187824	0.042858	2.238495	0.535372	0.535372	0.535372	0.535372	0.205882	0.238992	0.371133
N45	MNET-SS																																	
N46	PPC	1.41346	0.876705	0.812249	1.41346	0.876705	0.011707	0.818586	0.569507	0.871132	49.75372	0	0.988187	1.28097	1.439290	0.172709	0.120675	0.153328	0.182178	0.182178	828.1	519.4	0.218731	0.127102	0.133872	0.127102	1.857874	0.298104	0.302251	0.369779	0.298104	0.126288	0.133353	0.089764
N47	ZAPTRONIX																																	
N48	WETHLYS																																	
N49	FOSCHINI	3.103087	2.200818	2.004498	3.103087	2.200818	0	2.565088	0.869779	7.42777	157.2042	-1	1.203245	0.105419	1.628536	0.134391	0.084039	0.164736	0.136861	0.136861	61.2	61.2	0.038725	0.027134	0.043453	0.027134	1.353456	0.28118	0.321258	0.321258	0.28115	0.111571	0.42919	0.112102
N50	COMPCLEAR	4.630502	4.198823	4.198823	4.630502	4.198823	0.478751	0.626841	13.32177	11.00838	34.26077	0	0.787301	2.083822	6.998248	1.38037	0.275590	0.244034	0.258552	0.258552	11.23836	11.23836	0.008834	0.007307	0.004995	0.007307	1.222844	0.14075	0.14075	0.170867	0.14075	0	0.016417	0
N51	PALS	1.424177	0.780084	0.780084	1.424177	0.780084	8E-05	0.522001	0.461184	5.373455	57.14037	0	2.329521	11.75398	0.161337	0.957094	0.03765	0.257504	0.228399	0.228399	18.2	19.2	0.112002	0.042141	0.159008	0.042141	2.657772	0.602874	0.902674	0.902674	0.602874	0.344083	0.344083	0.25468
N52	CARGO	0.800066	0.78634	0.78634	0.806727	0.773014	0.112761	7.378516	90.10283	6.247800	-5.231947	0	1.325186	1.05977	2.830119	0.072144	0.03428	0.155680	0.097016	0.097016	42.9	24.6	0.197097	0.092595	0.246858	0.092595	2.135768	0.474579	0.479108	0.566364	0.470786	0.443424	0.444737	0.294023
N53	JASCO																																	
N54	PINNACLE	1.335804	0.688446	0.46283	1.335804	0.688446	0.034787	7.818822	4.750213	4.383978	41.50119	0	1.787117	0.032842	6.232191	2.883829	0.217374	0.447944	1.788467	1.788467	4.8	4.6	0.914831	0.133485	0.25081	0.133485	4.808408	0.715553	0.715553	0.715553	0.715553	0.004882	0.006236	0.344871
N55	LA-GROUP	1.356132	0.86372	0.863612	1.356132	0.86372	0.081087	2.380098	3.878299	3.816484	158.3053	0	1.077228	29.95453	2.888408	5.167819	0.130145	0.332949	0.350014	0.349572	34.08007	45.5	0.811424	0.004578	0.018061	0.004578	2.499801	0.580186	0.852529	1.04808	0.686185	0.014842	0.183247	-0.01052
N56	PUTCO	0.773142	0.558508	0.88613	0.773142	0.558508	0.008317	108.3188	20.08733	6.213157	-37.04288	0	1.505544	2.055808	2.516412	0.006233	0.002073	0.011531	0.005215	0.005218	4.8	19.7	0	0	0.101073	0	1.078708	0.348118	0.348118	0.352353	0.348118	0.348588	0.348589	0.651892
N57	GROUP-5	1.049082	0.786100	0.708332	1.049082	0.758188	0.127840	4.822179	9.674337	3.888083	18.52569	0	1.731208	0.88011	6.230280	0.029700	0.002895	0.072243	0.018035	0.018035	52.1	52.1	0.043036	0.011958	0.063508	0.011958	3.808611	0.701235	0.701235	0.788588	0.701235	0.344988	0.474713	0.124
N58	CONCOR	1.383923	0.647404	0.545128	1.383923	0.547404	0.132171	13.38629	4.894446	4.983556	25.08057	0	2.285147	17.18167	6.005800	0.038418	5.03372	0.225908	0.202521	0.202521	291.8	291.8	0.128576	0.048922	0.121993	0.048922	2.628237	0.805002	0.775156	0.785377	0.805002	0.193771	0.23036	1.295434
N59	CONTROL	1.900857	1.549688	1.549688	1.900857	1.549688	0.401611	6.603076	6.621223	5.380776	25.45568	0	1.300791	6.517086	2.333333	0.048774	0.047778	0.16208	0.111476	0.111476	9.9	9.9	0.063438	0.0468	0.143529	0.0468	1.782858	0.43405	0.43405	0.44695	0.43405	0.334950	0.428798	0.45239
N60	DCENTRIX																																	
N61	ENSERVE	1.421339	1.341732	1.218800	1.421339	1.341732	0.133213	5.885703	54.3786	6.480780	12.83759	0	1.209004	2.004812	2.872657	0.105625	0.072308	0.130088	0.207717	0.207717	26.5	22.1	0	0	0	0	2.370053	0.279278	0.388538	0.584838	0.279275	0.04962	0.080180	0.02678
N62	SOFTLINE																																	
N63	MONEYWB	3.84657	3.84657	3.831362	3.84657	3.84657	0.753775	3.341902	50	2.991078	-6.806251	0	0.299264	1.951965	0.428298	0.500821	0.546358	0.02875	0.234004	0.234004	2.2	2.2	0.066530	0.048403	0	0.048403	1.431172	0.266607	0.266607	0.266607	0.266607	0	0.030473	0
N64	CHEMSERVE	1.279668	0.66205	0.66205	1.279668	0.66205	0.038782	5.200347	5.016484	5.819626	50.16985	0	1.828298	18.05563	5.608502	0.121108	0.042382	0.213408	0.237868	0.237868	61.5	107.1	0.070254	0.02291	0.263831	0.02291	3.066517	0.612479	0.612479	0.697295	0.612479	0.211528	0.256285	0.38907
N65	DATATEC																																	
N66	WINHOLD	1.464036	0.920058	0.920058	1.464036	0.920058	0.067884	5.514448	7.72608	5.602873	50.53356	1	1.806733	5.042071	7.686377	0.036483	0.812576	0.087732	0.068884	0.068886	0.3	0.3	0.679464	0.150671	0.108095	0.150671	4.255404	0.580557	0.580557	0.80382	0.580557	0.415174	0.52046	0.81046
N67	PASDEC	1.489158	0.630842	0.630842	1.489158	0.630842	0.613452	9.504042	4.503743	7.201228	86.47686	0	1.438446	3.225241	3.038384	0.088051	0.872907	0.314004	0.221583	0.221583	5.8	5.8	0.324584	0.15362	0.524800	0.15362	2.11297	0.525645	0.548804	0.548804	0.525645	0.095472	0.724389	0.788573
N68	OMNIA	1.118291	0.721087	0.721087	1.118291	0.721087	0.018837	3.713617	9.45585	3.886188	80.87825	0	1.578823	5.336566	5.064486	0.084403	0.045848	0.345202	0.232188	0.232185	98.7	98.7	0.220395	0.088707	0.227758	0.088707	3.207781	0.884418	0.740323	0.831877	0.884418	0.43453	0.462026	0.530544
N69	ERP.COM																																	
N70	MTN	1.212886	1.158004	1.158004	1.212886	1.158004	0.151191	9.808767	94.07817	4.043514	-30.94857	0	9.823000	1.198283	1.408086	0.088292	0.071313	0.083049	0.108678	0.105878	6.4	6.4	0.364349	0.201778	0.214124	0.201778	1.805881	0.436753	0.855862	1.132943	0.436753	0.178975	0.271389	0.087541
N71	MATHOMO	1.383358	0.868088	0.886760	1.383358	0.868088	0	3.232246	5.438894	5.920954	118.367	0	1.400431	19.25271	4.88811	0.107851	-0.075576	-0.008872	-0.362581	-0.35258	-29.9	-3.9	0.265703	0.078782	0.452080	0.878782	3.331197	0.888808	0.757532	0.773334	0.888808	0.458425	0.507881	0.850428
N72	MASONITE	2.140723	1.211282	1.187003	2.80454	1.15506	0.008742	8.788836	9.833198	7.278724	87.95319	0	1.443036	2.882158	2.267836	0.84727	0.030286	0.085968	0.088881	0.088881	187.7735	108	0.10534	0.087028	0.094315	0.087028	1.571574	0.301086	0.301086	0.32844	0.305098	0.324814	0.38431	0.356418
N73	ADONIS	2.718009	1.103818	1.09841	2.718009	1.103818	0.047038	4.868248	2.002279	5.508024	137.448	0	1.123882	3.403841	1.733583	0.114176	0.044955	0.131748	0.077834	0.077834	20.9	20.9	0.109088	0.070883	0.197165	0.088871	1.542880	0.316381	0.316381	0.319381	0.316358	0.368585	0.604545	0.458477
N74	APLITEC																																	
N75	GRINTEK																																	
N76	DELTA	1.818902	0.872102	0.872102	1.818902	0.872102	0.146980	5.769708	3.94238	6.923479	100.7146	0	1.18875	4.038591	2.378548	0.189467	0.117358	0.2803	0.278788	0.278788	225.5	232.9	0	0	0.114830	0	1.98495	0.408388	0.408388	0.499418	0.408388	0.100818	0.134911	0.112442
N77	AMAFE	1.528046	0.957308	0.582129	1.528049	0.957308	0.000120	5.515798	5.430412	12.67372	105.0463	0	1.77083	13.93738	7.946875	0.030249	0.837348	0.184244	0.298777	0.298777	0	0	0.924074	0.205887	0.040082	0.205887	4.488049	0.777188	0.777188	0.854878	0.777188	0	0.310883	0
N78	STELLA																																	
N79	VALUE	1.259245	1.123264	1.123264	1.259245	1.123264	0.001454	4.385384	28.48824	8.779682	41.77182	0	1.0129	1.49188	1.505796	0.428578	0.283384	0.338373	0.398573	0.398573	12.7	12.8	0.054082	0.036379	0.040049	0.036379	1.40862	0.243328	0.243328	0.337087	0.243328	0	0	0
N80	REBSERV	1.887445	1.374305	1.374305	1.887445	1.374305	0.33337	10.16307	15.65888	5.973311	-5.218881	0	1.629529	32.42128	6.768586	0.058056	0.84587	0.147174	0.284851	0.284851	33.1	33.1	0.973851	0.308839	0.082047	0.308838	3.153583	0.682003	0.890127	0.782319	0.682003	0.116445	0.324014	0.073206
N81	MT-EAGLE																																	
N82	AF-&-OVER	5.334258	2.864678	2.884678	5.334258	2.864678	0.225234	0.014678	3.427162	10.91829	132.6117	1	1.307938	0.398801	2.880508	0.075484	0.038585	0.058813	0.108438	0.107834	78	78	0.007082	0.003182	0	0.003102	2.288435	0.157838	0.15881	0.204272	0.157838	0	0.305528	0
	Max	6.297189	6.207497	5.148958	6.297189	6.207497	0.755778	183.3185	6540.2	23.89828	278.9187	1	2.847474	1390.352	18.84315	12.34481	2.231033	2.035178	17.53211	17.53211	520.1	518.4	7.420146	0.911812	0.452080	0.811812	18.77831	0.880547	0.880547	4.958928	0.880547	0.954283	0.88434	5.890957
	Average	2.028941	1.530174	1.474535	2.028086	1.52929	0.14787	7.639756	133.8049	7.220543	67.30572	0.065574	1.321374	33.81037	3.524211	0.488883	0.158419	0.22472	0.591238	0.591208	58.93504	82.77858	0.28834	0.095459	0.118252	0.088112	2.405288	0.458328	0.401855	0.684451	0.458338	0.174273	0.288891	0.327886

All Year 3 and Other Potential Variables Selected [Format: Variable Name (Year Prior to Failure)]																																Target Label	Short Name	Ref Code
C4(3)	C5(3)	C6(3)	C7(3)	C8(3)	C9(3)	C10(3)	C11(3)	C12(3)	M1(3)	M2(3)	M3(3)	M4(3)	M5(3)	M6(3)	G1(3)	G2(3)	G3(3)	G4(3)	G5(3)	G6(3)	N1(3)	N2(3)	N3(3)	R1	R2	R3	R4	R5	R6	N4	N6	T		
31.41011	0.510454	0.236868	0.311202	4.484331	0.989262	0.109937	0.373932	0.999453	0.081292	0.029444	0.13251	3.086777	1.524416	7.91842	8.339311	6.286836	0.221183	0.824809	0.167371	0.082862	0.286901	0	0	0.136977	0.05319	2.911191	0.128863	0.004223	-27.8469	41	56	1	CAXTON	N1
5.100587	0.708485	0.436238	0.342546	0.426232	0.114858	0.02195	0.635825	0.082967	0.120703	0.914188	0.058844	6.648721	3.684193	12.56732	5.995089	6.389926	0.639219	0.950864	0.221744	0.42526	0.910283	0	0	0.836698	0.058645	-59.9626	1.316635	0.084376	11.09925	5	86	1	TRUWTHS	N2
1.762202	0.260337	0.176454	0.033735	0.805057	-0.10336	0.130045	2.11916	0	0.188665	0.03008	0.106212	3.546433	6.215456	11.15767	5.133044	5.486743	0.085342	0.715482	0.447266	0	0.286594	0	0	0.186659	0.160866	1.1001	0.006134	0.121399	0.061664	14	37	1	INVICTA	N3
																								1.579104	0.250186	-1.60442	4.166829	0.162431	-0.17642	6	6	1	AMB	N4
0.788502	1.608431	0.18999	0.12752	-0.66646	0.368335	0.718716	3.505848	0	0.298911	0	0.9172	15.50849	8.86413	5.579008	5.448737	6.195424	0.186323	1	9.141175	0.078749	0.286664	0	0	0.198501	0.397563	-0.09533	0.185986	5.1641	1.153579	5	34	1	MGX	N5
205.1	0.004139	0.264274	0.217843	4.244166	0.340087	0.00929	0.407282	0.573626	0.127512	0.015369	0.051434	3.28268	2.540169	12.45892	5.881409	6.035621	0.227229	0.698972	0.090889	0.372033	0.104993	0	0	0.482994	0.17526	1.205895	0.379209	0.126399	-2.15429	3	70	1	MRPRICE	N6
																								1.027184	0.097795	4.420616	0.207646	0.068323	-3.73651	7	7	1	HOWDEN	N7
0.766993	1.033652	0.399679	0.394267	-0.76712	0.303468	1.635194	23.48333	0	0.053029	0	0.043478	8.70435	12.25896	14.43351	4.252319	4.96601	0.219376	1	0.655443	0	0.379929	0	0	0.718999	0.246798	-1.1519	0.971955	0.02962	19.62113	5	5	1	INFOWAVE	N8
1.765117	0.669814	0.326321	0.254267	-0.09652	0.085119	0	0.191795	0	0.094919	0.057143	0.295652	0.266632	0.424478	3.931124	6.221448	5.236753	0.142144	0.865708	0.109652	0.148311	0.221004	0	0	0.166609	0.347958	1.093477	0.167177	0.352317	0.530061	8	35	1	BEARDEL	N9
0.530022	0.004634	0.286881	0.344292	-0.99795	0.702194	0.362713	13.39052	0	0.33935	0	0.04	11.16103	13.59362	3.554534	4.389639	6.176265	0.240117	1	0.605671	6.20915	0.646077	0	0	0.409203	0.438566	-0.33471	0.134999	0.248146	-0.40946	5	5	1	EOH	N10
1.369907	0.916011	0.210625	0.065768	0.606602	0.032986	0	0.201572	0	0.630093	0.614795	0.092087	2.298963	1.331794	9.995358	6.474262	5.369828	0.223822	0.83931	0.192614	0	0.466963	0	0	0.2359	0.282575	0.907931	0.095369	0.211926	0.275025	16	20	1	NUWORLD	N11
																								0.423696	0.060102	5.28617	0.813676	0.127339	3.883451	13	17	1	LABAT	N12
																								3.298063	0.163112	-68.8078	5.486007	0.180424	7.604367	8	11	1	AWETHU	N13
2.842813	0.310224	0.364857	0.32732	0.916282	0.463671	0	0.242055	0.711871	0.421106	0.917423	0.056092	2.933699	3.546464	13.87123	6.186128	6.366873	0.14126	0.682704	0.132545	0.492128	0.012998	0	0	0.235172	0.125082	1.606024	0.034302	0.062178	-0.61493	7	7	1	NUCLICKS	N14
																								0.880842	0.087651	-11.3887	1.417617	0.019766	36.31863	66	86	1	REX-TRUE	N15
1.208356	2.801902	0.156832	0.91741	-0.63626	-0.19206	0	0.457703	0	0.072632	0.04965	0.112162	0.90289	2.400663	8.827465	4.265242	4.102226	0.202893	0.579759	0.175105	0	0.67	0	0	0.162527	0.070326	-0.80895	0.213442	0.009363	19.93809	18	43	1	SPANJAARD	N16
0.831646	0.127506	0.212303	0.146333	0.093123	0.065302	0	0.290957	0	0.094818	0.1875	0.575	0.275056	0.416781	0.261444	4.874777	4.081512	-0.28776	0.674081	0.159313	0.03404	0.107102	0	0	0.456432	0.182727	1.770066	0.574932	0.132766	6.611633	23	34	1	INMINS	N17
17.56346	0.373079	0.654758	0.670172	4.146077	0.90386	0.048421	0.672271	0	0.06591	0	0.029762	36.27845	2.947466	48.82214	5.507172	5.903578	0.084972	1	0.081389	4.916421	0.113784	0	0	1.357335	0.170127	72.63432	1.102783	0.218173	19.41206	13	17	1	KGMEDIA	N18
0.704943	1.660487	0.252029	0.191122	-1.26614	-0.3399	0.752114	2.922411	0	0.259996	0.007333	0.246	10.86218	6.2635	10.395	5.190068	5.78547	0.052186	0.675446	0.184692	0.379059	0.254155	0	0	0.313621	0.514436	0.173055	0.217711	0.232336	-0.67863	8	8	1	SASANI	N19
0.72502	3.70184	0.143999	-0.10507	0	-0.13611	1.88E-06	0.239874	0	0.604471	0.030544	0.092336	3.995652	1.840627	6.661695	6.252246	6.29861	0.549983	0.689178	0.113745	0	0.081748	1	1	0.084094	0.131259	0.646914	0.006486	0.026601	0.221618	34	53	1	ELLERINE	N20
2.066452	0.035039	0.665478	0.938714	-0.26371	1.277066	0.582654	1.496312	0	0.316032	0.071429	0.342857	2.506578	0.634095	2.426791	4.721473	4.420266	0.110295	0.755483	0.18368	0	0.53683	0	0	7.215361	0.182044	-1.77825	-2.34499	0.104627	-7.80777	5	6	1	GLOVIL	N21
3.166501	2.499935	0.226863	0.043269	0.369677	-0.33647	0.000442	1.232954	0	0.134578	0.104651	0.061626	0.448166	0.605021	3.440092	5.630517	4.956331	0.227764	0.498711	0.08996	0.00705	0.203716	0	0	0.137743	0.081394	-1.15222	0.144285	0.046653	-3.55773	18	28	1	SEARMAN	N22
1.155575	0.130604	0.585415	0.339751	0.50134	2.750839	0.08136	3.203611	0	0.007167	0.014483	0.044203	11.90466	3.916576	11.09075	5.026663	5.489331	0.078534	0.701051	0.134903	0	0.499392	0	0	0.232646	0.476214	0.149441	0.161175	0.160866	-0.76232	5	10	1	UCS	N23
																								1.163318	0.166029	4.636607	0.31686	0.074644	-3.27666	6	14	1	CONNECT	N24
1.036191	2.404156	0.025653	-0.06748	-0.29076	0.726006	0.227952	1.472372	0	0.353178	0.009342	0.025482	11.30944	6.266675	33.80958	6.185321	6.859162	0.082623	1	0.145711	0.182076	0.179474	0	0	1.007786	0.665619	1.263043	0.667659	0.40642	-0.79524	9	9	1	COMPAREX	N25
0.497932	0.720931	0.425194	0.227603	0.359632	0.287866	0.117984	0.127469	0	0.034521	0.914616	0.03919	6.401421	5.775349	15.86865	6.462013	5.943045	0.187535	0.590098	0.130508	0.303888	0.255429	0	0	0.066622	0.134938	0.606674	0.0682	0.102033	0.651617	15	21	1	CITYLDG	N26
1.517792	1.262467	0.345791	0.082637	-0.43491	0.13521	0	0.112705	0	0.198924	0.081301	0.764228	0.993515	0.566004	3.243228	5.152175	4.604569	0.434402	0.802309	0.354077	0.116375	0.578093	1	0	0.649473	0.083795	-1.39678	8.863673	0.078627	26.64758	15	62	1	TRNPACO	N27
1.242017	0.83969	0.33866	0.191473	0.00416	0.284931	0	0.111678	0	0.219768	0.103226	0.32	0.857447	0.912586	4.582054	5.243207	4.764846	0.346105	0.689519	0.135376	0	0.026842	0	0	0.277302	0.125163	1.922076	0.627896	0.911677	2.416931	8	9	1	ARGENT	N28
1.477674	2.62921	0.226829	0.095	-0.35349	0.3265	0	0.120923	0	0.029754	0.674074	0.284815	0.274621	0.343711	5.487641	4.403892	3.585179	0.283957	0.721022	0.138671	0	0.675086	0	0	0.113835	0.083748	3.130443	0.144441	0.103305	1.361293	34	39	1	NICTUS	N29
369.0984	0.004986	0.893701	0.987837	354.8675	8.946349	0.907477	5.622272	0	0	0	0.104	7.461103	6.320984	0.059266	5.496264	6.033969	0.089098	1	0.774612	0.117277	0.242272	0	0	1.354144	0.716217	-0.14964	0.211894	0.270769	-0.54983	4	18	1	AST GROUP	N30
0.795722	1.968866	0.258295	0.239173	0.194266	-0.03135	0	1.923079	0	0.182898	0	-2.42667	0.221664	-1.18291	-2.38656	5.42747	4.696058	-0.41061	1	2.777426	0	0.166761	1	0	-2.86491	0.123463	-16.8987	2.648178	0.174185	-15.276	4	7	1	SEKUNJALO	N31
35.23626	0.605994	0.369932	0.342941	17.65621	0.141149	0.061447	1.052291	3.626946	0.14266	0.0151	0.037337	6.201021	3.114267	15.2178	6.725817	7.817926	0.14136	0.909757	0.102249	0.126436	0.073209	0	0	0.334839	0.166949	2.545451	0.129427	0.16987	1.217527	14	67	1	IMPERIAL	N32
0.303367	2.262682	0.203533	0.050398	-0.05312	0.063194	0	1.333333	0	0.178108	0.024528	0.118868	4.77465	2.208023	74.32873	6.976998	6.154608	0.009932	0.014637	0.003613	0	0.32933	0	0	0.150836	0.30592	0.010697	0.17668	0.330377	0.592765	34	35	1	MOBILE	N33
18.22436	0.093863	0.094336	0.6665	-8.76923	0.341374	1.16999	36.48856	0	0.22536	0	0.043662	6.098088	6.361093	2.003276	4.716621	5.263436	0.130084	1	0.441811	0	0.173657	0	0	1.785261	0.238033	-2.39047	22.114	0.25647	-4.20271	5	17	1	MAXTEC	N34
8.895467	0.58666	0.560121	0.356143	0.808972	-0.0032	0.161322	1.369432	0	0.527006	0	0.098174	2.998476	23.98097	0.467626	6.098165	6.253304	0.046907	1	17.63211	0.693334	0.123979	0	0	1.246914	0.087076	-4.80733	0.412808	0.012534	46.10026	1	10	1	PRIMA	N35
																								0.289518	0.143874	2.105193	0.038536	0.108845	-0.31813	4	16	1	MUSTEK	N36
																								1.274642	0.315733	12.86546	1.074134	0.164495	23.99684	18	35	1	ADCORP	N37
0.746464	26.67564	0.024859	-0.19767	2.471688	8.227099	0	-4.21369	0	0.058461	0	-0.07642	3.329764	1.078448	2.957748	5.377762	5.240732	-0.00264	1	-0.08257	140.7134	0.846844	0	0	9.592834	0.630615	0.736112	-2.49406	0.543926	2.297336	12	14	1	GOLDREEF	N38
-7.66542	-0.79337	-0.16075	-0.14354	2.144652	0.176547	0.298058	1.266715	0	0.045107	0	0.173176	1.793476	2.917014	1.834136	5.203479	6.179307	0.169008	0.940285	0.918368	0.116519	0.637849	1	1	3.769	0.253605	3.630847	-2.75301	0.22573	14.93936	4	28	1	HERCOL	N39
																								0.224685	0.133167	2.126764	0.080718	0.070086	1.276849	8	43	1	STEERS	N40
-1.56209	-0.25598	-0.4569	-0.36209	-0.98167	0.933932	0.732927	12.45269	0	0.208142	0	0.043187	6.158404	3.059036	20.90834	5.198825	5.543708	0.086065	1	0.118293	0	0.411993	0	0	0.736776	0.409792	2.993026	0.4993	0.226716	2.858458	6	6	1	BUSBY	N41
-0.01219	0	-0.01111	0.003748	-1.00006	-0.21769	1.260013	31.54117	0	0.313682	0	0.008	16.22951	123.9859	119.5265	4.931397	6.083742	0.102298	1	1.010811	78.82764	0.471926	0	0	0.860306	0.672364	5.63575	0.215525	0.447954	-0.28626	6	6	1	FRONTRNGE	N42
-7.32056	-1.35326	-0.12452	-0.12702	-0.26829	0.152241	1.116895	23.15658	0.570416	0.095218	0.016529	0.066596	6.357125	2.941141	16.60347	5.376639	5.856099	0.149748	0.805144	0.13583	0	0.158882	0	0	1.233757	0.286829	7.737193	1.190536	0.040275	161.573	8	8	1	BRANDCO	N43
2.09926	1.204039	0.16727	0.137866	-0.63922	0.307723	0.109769	1.297472	0	0.864025	0	0.91465	2.882822	3.095326	8.874861	5.445101	6.595862	0.08201	1	0.222519	0	0.105083	0	0	0.626263	0.201897	30.85946	0.064558	0.208303	0.340045	15	26	1	SPESCOM	N44
																								1.129057	0.034402	-7.88099	1.007145	0.025978	-22.3146	13	18	1	MNET-MS	N45
1.58037	0.666967	0.67637	0.460154	-0.43206	-0.26986	0.009626	0.340456	0.689931	0.160191	0.101404	0.162278	2.304312	1.131261	4.973832	6.344864	6.178649	0.380208	0.352554	0.084226	0.09308		0	0	0.192846	0.039675	0.370536	0.239376	0.041714	-7.11164	93	111	1	FPC	N46
																								4.241457	0.370734	3.86526	-21.0171	0.566765	4.992565	6	6	1	ZAPTRONIX	N47
																								0.213231	0.274886	0.872649	0.167986	0.176091	1.094336	5	30	1	WETHLYS	N48
5.931217	0.269997	0.616273	0.395465	1.823633	0.102604	0	0.819084	0.60191	0.214125	0	0.076	5.156866	1.823331	7.622454	6.320063	6.446465	0.634169	0.8125	0.1112	0	0.06121	0	0	0.276083	0.066537	-2.29677	0.346259	0.083943	-3.11693	62	66	1	FOSCHINI	N49
37.65675	0.916117	1.955009	1.781395	22.99575	6.245203	1.661726	3.708447	0	0.2239	0	0.03468	37.67366	6.477168	2.674623	4.4608	5.190745	0.109267	0.771642	0.189509	0.046682	0.393276	1	0	0.752991	0.170053	-1.82596	0.258364	0.106652	-1.69382	5	19	1	COMPCLEAR	N50
4.337607	0.673153	0.303302	0.202226	0.16463	0.191981	0	0.179237	0	0.038	0.070833	0.4	0.35694	0.574369	5.014061	4.346568	3.681241	0.174808	0.82284	0.186704	0	0.6188	0	0	0.423966	0.178123	3.544824	0.097611	0.089205	1.10121	16	66	1	PALS	N51
1.860446	1.449086	0.362873	0.272877	0.083319	0.0614	0	0.260672	0	0.0315	0.075342	0.293151	0.326912	0.331254	3.320872	5.274777	4.466383	0.431163	0.74275	0.072059	0.128641	0.6725	0	0	0.077602	0.106064	-0.42076	0.109614	0.142133	-0.6436	15	47	1	CARGO	N52
																								0.316576	0.467163	0.819103	0.227194	0.506171	0.341254	16	27	1	JASCO	N53
0.520465	3.69765	0.097089	0.04676	-0.80447	-0.11465	0.195407	0.678108	0	0	0	0.383333	0.150021	0.494469	0.068164	5.130749	4.161543	-0.15304	1	1.789487	0	0.231697	0	0	1.672493	0.114179	6.412651	0.934744	0.015279	27.22229	18	17	1	PINNACLE	N54
-59.4296	-0.04036	-0.45366	-0.43362	-98.662	-0.40455	1.154939	35.79926	0	0.167323	0.91736	0.05471	4.278849	6.397639	16.69917	5.093863	5.802533	0.192665	0.653639	0.228783	0.411434	0.412019	0	0	1.199477	0.582265	4.832262	1.069386	0.582325	4.79069	3	67	1	LA-GROUP	N55
420	1.189291	0.246539	0.173494	350	0.039843	0	0.152194	0	0.133434	0.941494	0.020332	0.441124	0.258095	9.408466	5.620245	4.804032	0.137478	-0.4051	-0.00211	0.008512	0.111193	0	0	0.78304	0.006814	417.6974	0.353047	0.000979	286.9146	15	15	1	PUTCO	N56
12.55516	0.358376	0.214107	0.175396	0.302421	0.16209	0.189097	1.098586	0	0.058296	0.026249	0.073566	0.629761	1.338912	11.93346	5.148808	5.716752	0.055394	0.584299	0.010536	0.347578	0.001726	0	0	0.312463	0.093151	2.374143	0.286954	0.083776	3.703098	28	34	1	GROUP-5	N57
0.366547	6.803044	0.02964	-0.01166	0.371367	-0.12174	0	0.633715	0	0.253671	0.042984	0.147646	0.082582	1.371934	6.839365	5.703861	5.421426	0.260397	0.714478	0.144696	0.099333	0.002622	0	0	-32.0289	0.041931	33.95116	-1.66549	0.016553	191.4986	22	55	1	CONCOR	N58
1.841102	1.579946	0.209295	0.233026	0.94117	0.352703	0	-0.16601	0	0.962148	0.0625	0.158125	0.697258	0.899603	7.514225	5.167465	4.763056	0.16676	0.672196	0.974935	0.06396	0.358204	0	0	0.280093	0.337861	-0.2316	0.249646	0.241949	-0.95517	18	39	1	CONTROL	N59
																								0.492141	0.423616	-0.48029	0.720005	0.320541	-1.16107	5	31	1	DCENTRIX	N60
420	0	0.847234	0.736332	350	0.285754	0	0.18503	0	0.276536	0.112903	0.330645	0.946216	0.629863	2.811651	6.436197	4.861601	0.150407	0.657645	0.138604	0.325176	0.1619	1	0	0.156359	0.076261	1.626767	0.199726	0.003622	-64.3734	7	9	1	ENSERVE	N61
																								0.506746	0.191515	10.7313	0.154637	0.013398	12.83584	13	15	1	SOFTLINE	N62
2.152	0	0.375279	0.566127	-1.8364	2.457031	5.48673	32.67977	0	0.21955	0	0.021782	22.54464	8.602151	29.00791	4.126496	4.907411	0.117236	1	0.234004	0	0.40605	1	0	0.704977	0.525027	2.556365	0.049802	0.171672	-0.24621	4	5	1	MONEYWB	N63
5.808386	2.05406	0.20971	0.915917	0.031915	-0.28632	0.000975	0.682308	0	0.126663	0.034636	0.099289	1.141621	2.14464	4.013717	5.914132	5.75862	0.367166	0.84476	0.153239	0.039544	0.001431	0	0	0.090866	0.059041	-0.47099	0.091678	0.063452	1.105404	36	67	1	CHEMSERVE	N64
																								0.92208	1.331437	-0.22128	0.799042	1.069236	0.419074	9	9	1	DATATEC	N65
0.259732	4.750126	0.070344	-0.01212	-0.28684	-0.28123	0.002696	0.483083	0	0.078462	0.060714	0.225	0.163894	0.430619	7.876998	5.260974	4.286327	-0.19147	0.864253	0.08356	0.039806	0.195229	0	0	0.342099	0.202618	1.640766	0.362903	0.160084	2.605367	57	68	1	WINHOLD	N66
0.762042	2.775968	0.222707	0.10962	0.073617	0.404209	0	-0.08255	0	0.119762	0	0.112	0.251664	0.279716	3.444863	4.987763	4.089587	0.000571	1	0.221693	0	0.609755	0	0	13.88643	0.116872	9.081134	-1.44369	0.136726	657.4406	15	18	1	PAEDEC	N67
1.653565	2.004731	0.165996	0.915203	-0.51494	0.079647	0	0.248733	0	0.083466	0.049686	0.094269	1.122234	2.464027	8.646616	5.722943	5.608366	0.168257	0.401664	0.107191	0.308795	0.208423	0	0	0.254556	0.134699	2.455824	0.072257	0.002137	35.56367	46	50	1	OMNIA	N68
																								0.965136	0.434876	-0.46974	0.326767	0.287871	1.166046	6	8	1	ERP.COM	N69
0.862677	1.202226	0.407795	0.365101	-0.70607	-0.43018	0	0.548627	0	0.081742	0.003663	0.01647	6.772451	6.676847	23.34691	5.647702	6.429577	-0.11664	0.76093	0.060642	0.302487	0.004999	0	0	1.431926	0.469724	20.26966	1.169674	0.376913	12.31353	5	8	1	MTN	N70
1.332149	4.254762	0.151934	0.030176	3.832741	0.134114	0.148774	0.475646	0	0.474147	0	-1.15	0.134089	0.312586	2.292234	5.303436	4.275808	-0.22726	1	-0.35256	0	0.19142	1	0	-5.60263	0.16876	36.16058	-0.74017	0.167356	7.190545	7	8	1	MATHOMO	N71
0.929459	1.51369	0.206918	0.046954	-0.62689	-0.21934	0	0.131502	0	0.126777	0.049637	0.260636	0.678121	0.274026	8.970215	5.228866	4.467125	0.899603	0.985117	0.058298	0.019581	8.000087	0	0	0.631585	0.251699	-14.9549	0.666634	0.619645	-22.6182	51	61	1	MASONITE	N72
1.805304	1.007047	0.333191	0.092636	0.436637	0.183778	0	0.171471	0	0.036364	0.034091	0.11676	1.36174	0.658205	3.699653	4.168401	3.795129	0.498396	0.712551	0.095632	0	0.783921	0	0	1.042196	0.167774	4.615736	1.845193	0.230472	4.039878	34	51	1	ADONIS	N73
																								0.959507	0.662	0.612666	0.5636	0.406835	1.642093	6	6	1	APLITEC	N74
																								0.375236	0.417676	0.913316	0.496353	0.563276	0.6266	14	39	1	GRINTEK	N75
420	0.49697	0.567624	0.33805	350	0.131401	0	0.193264	0	0.117276	0.036169	0.089626	3.732832	3.016987	7.146507	6.911419	6.092376	0.341505	0.640799	0.176648	0.14194	0.911492	1	0	0.276889	0.074492	4.931678	8.039717	0.914298	2.829134	31	46	1	DELTA	N76
0.341392	0.666637	0.090444	0.106612	0.353263	0.182001	0	0.28623	0	0	0	0.05	1.701678	8.936572	27.80669	4.935427	5.056905	0	1	0.298777	0.049046	0.463158	0	0	0.417096	0.193196	3.151142	0.249961	0.096928	2.654481	6	6	1	AMAPS	N77
																								0.510767	0.602614	0.651863	0.416763	0.314032	3.473601	5	9	1	STELLA	N78
9.5166	0.133007	1.422839	1.380686	0.2916	0.233322	0.030711	0.614431	0	0.085925	0	0.230909	3.414151	1.235013	1.509046	5.136116	5.057586	0.281481	1	0.398573	0.136095	0.549996	0	0	0.294543	0.40041	-0.33647	0.368788	0.232251	-1.05262	4	6	1	VALUE	N79
0.654136	0.307127	0.298219	0.279206	-0.6665	0.355737	0.261673	7.137954	0	0.272537	0.008	0.0331	4.107467	8.934176	26.71344	5.677193	6.124553	0.103557	0.987351	0.261303	0.002609	0.12943	0	0	1.266606	0.666168	8.917392	0.975451	0.5654	3.437704	2	8	1	REBSERV	N80
																								0.163046	0.372822	-8.245107	0.123201	0.40689	0.369346	21	21	1	MT-EAGLE	N81
43.94909	0	0.864799	0.649523	0	0.361697	-0.00044	0.443697	0.49983	0.024757	0.047101	0.137691	2.236336	0.800044	3.94441	5.248741	4.795402	0.376791	0.697198	0.098339	0.286723	1	0	0	0.098746	0.110158	-0.71477	0.098752	0.079266	-0.97336	55	58	1	AF-&-OVER	N82
420	26.67564	1.955009	1.781395	354.8675	8.946349	5.48673	36.48856	3.626946	0.962148	0.1875	0.764228	38.27845	123.9859	119.5265	6.725817	7.917926	0.899603	1	17.63211	140.7134	1	1	1	13.88643	1.331437	417.6974	22.114	1.069236	657.4406	93	111		Max	
33.74042	1.674079	0.319431	0.239164	22.17361	0.440734	0.326909	4.26266	0.136966	0.196099	0.030472	0.073339	6.184348	6.429027	12.38659	5.371069	5.327465	0.191239	0.780329	0.549109	3.865157	0.300554	0.147541	0.147541	0.544072	0.262036	7.068246	0.378779	0.183828	16.95809	16.23171	28.34146		Average	

APPENDIX G – RATIO CORRELATIONS

This appendix consists of the following components:

- **Appendix G.1**: A correlation matrix of all features.
- **Appendix G.2**: A t-test of all the correlations in Appendix G.1. Highlighted values are significant at either the 1% or 5% level (see Appendix G.2.)

Appendix G.1. A correlation matrix of all features.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	E1	E2	E3	P1	P2	P3	P4	P5	P6	P7
L1	1.00																				
L2	0.98	1.00																			
L3	0.59	0.59	1.00																		
L4	0.97	0.95	0.56	1.00																	
L5	0.97	0.99	0.58	0.98	1.00																
L6	0.18	0.20	0.53	0.18	0.19	1.00															
L7	-0.04	-0.05	-0.13	-0.04	-0.05	-0.05	1.00														
L8	-0.06	0.00	-0.02	-0.07	0.00	0.14	-0.02	1.00													
L9	0.14	0.14	0.33	0.19	0.17	0.15	8.10	0.08	1.00												
L10	0.09	0.06	8.18	0.09	0.07	0.02	0.00	-0.05	0.07	1.00											
L11	0.01	-0.02	-0.03	0.01	-0.02	0.04	0.01	-0.03	-0.07	0.00	1.00										
E1	-0.11	-0.11	-0.10	-0.12	-0.12	0.00	0.28	0.02	0.25	0.04	0.03	1.00									
E2	-0.03	-0.01	-0.08	-0.81	-0.01	-0.06	8.15	-0.02	0.11	0.01	-0.01	-0.01	1.00								
E3	0.00	0.00	-0.01	-0.01	0.00	0.00	8.01	-0.01	8.01	0.00	8.00	0.02	0.00	1.00							
P1	0.12	0.12	0.26	0.12	0.12	0.08	0.09	-0.01	0.03	0.69	0.00	8.00	0.02	0.00	1.00						
P2	0.03	0.02	0.04	0.03	8.02	0.04	0.02	-0.01	0.04	0.14	0.01	0.08	0.01	0.00	0.48	1.00					
P3	-0.05	-0.05	-0.08	-0.05	-0.05	-0.03	0.59	8.06	-0.04	8.00	-0.01	0.09	-0.02	-0.04	-0.08	0.04	1.00				
P4	0.02	0.01	0.03	8.02	0.01	0.01	0.01	0.05	0.02	0.04	0.01	0.01	0.01	-0.75	0.03	0.01	0.05	1.00			
P5	0.02	0.01	0.03	0.02	0.01	8.01	8.01	0.05	0.02	0.04	0.01	0.01	8.01	-0.75	0.03	8.01	0.05	1.00	1.00		
P6	0.05	0.04	0.13	0.02	0.03	0.17	-0.02	-0.04	-0.05	0.01	-0.06	-0.04	-0.04	0.01	8.08	8.02	-0.04	0.02	0.02	1.00	
P7	0.04	0.03	0.13	0.01	0.01	0.11	-0.04	-0.04	-0.05	0.02	-0.09	-0.07	-0.04	0.01	0.03	0.02	-0.05	0.02	0.02	0.77	1.00
S1	-0.02	-0.01	-0.03	-0.02	-0.01	-0.02	0.00	0.00	-0.01	-0.01	-0.01	0.01	0.00	0.82	0.00	0.00	-0.02	-0.88	-0.86	0.00	-0.01
S2	-0.06	-0.06	-0.08	-0.07	-0.08	-0.05	0.00	0.02	-0.05	-0.02	-0.04	-0.03	-0.01	8.02	-0.28	-0.03	0.34	-0.03	-0.03	-0.08	-0.07
S3	-0.28	-0.19	-0.31	-0.21	-0.28	-0.39	8.01	-0.09	-0.06	0.05	-0.03	-0.06	-0.04	0.02	-0.04	8.04	0.14	-0.11	-0.11	-0.24	-0.19
S4	-0.08	-0.05	-0.08	-0.07	-0.06	-0.05	0.00	8.02	-0.05	-0.02	-0.04	-0.03	-0.01	0.02	-0.20	-0.03	0.34	-0.03	-0.03	-0.08	-0.07
S5	-0.01	-0.01	-0.02	-0.01	-0.01	0.01	0.08	8.00	-0.01	-0.02	8.00	-0.01	0.00	8.97	-0.01	8.00	-0.04	-0.78	-0.78	0.02	8.01
S6	-0.21	-0.19	-0.26	-0.17	-0.18	-0.12	8.04	0.08	-0.09	-0.08	-0.02	0.12	-0.04	0.00	-0.22	-0.01	0.41	-0.02	-0.02	-0.17	-0.18
S7	-0.22	-0.19	-0.27	-0.18	-0.18	-0.10	0.04	0.10	-0.07	-0.03	-0.06	0.12	-0.04	0.00	-0.22	8.00	0.40	-0.02	-0.02	-0.18	-0.19
S8	-0.20	-0.17	-0.25	-0.17	-0.15	-0.09	8.69	8.09	-0.08	-0.04	-0.04	8.11	-0.05	0.00	-0.22	0.00	8.43	-0.04	-0.04	-0.14	-0.18
S9	-0.21	-0.18	-0.27	-0.22	-0.19	-0.10	0.05	8.09	-0.14	-0.04	0.00	0.13	-0.05	0.00	-0.22	-0.01	8.41	-0.02	-0.02	-0.15	-0.15
C1	-0.03	-0.03	-0.06	-0.04	-0.04	-0.08	-0.01	0.01	-0.01	0.01	-0.02	-0.04	8.02	8.05	8.00	0.00	-0.02	0.01	0.01	0.00	8.01
C2	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.11	-0.01	0.00	0.00	8.07	-0.01	0.02	0.11	0.01	0.00	-0.08	-0.01	-0.01	8.02	0.02
C3	0.00	-0.01	-0.01	0.00	-0.01	-0.03	-0.02	-0.01	8.03	0.01	0.04	8.02	0.04	8.02	0.00	0.00	8.13	8.81	0.01	-0.08	-0.05
C4	0.08	0.07	0.05	0.06	0.07	0.19	0.04	0.03	-0.01	-0.03	8.01	0.08	0.00	8.00	-0.02	-0.04	0.09	0.02	0.02	0.09	0.08
C5	0.00	-0.01	-0.02	-0.01	-0.02	-0.02	-0.02	-0.04	0.04	0.05	8.02	0.00	0.04	8.02	0.01	8.01	0.16	0.01	8.01	-0.09	-0.07
C6	-0.07	-0.08	-0.13	-0.07	-0.09	0.11	-0.01	-0.07	-0.03	0.05	0.01	-0.05	-0.03	0.00	8.27	0.78	0.02	0.05	0.05	0.14	0.14
C7	0.13	0.12	-0.08	0.13	0.12	0.15	-0.08	-0.06	-0.08	0.04	0.01	-0.20	-0.03	-0.02	0.28	0.69	0.08	8.08	0.08	0.18	0.12
C8	0.19	0.22	0.28	0.19	0.22	0.17	0.01	8.10	8.00	-0.04	-0.04	8.00	0.08	0.00	-0.08	-0.02	0.02	8.02	0.02	-0.02	-0.02
C9	0.28	0.29	0.08	0.27	0.29	0.08	0.01	-0.05	0.04	-0.04	-0.02	-0.01	0.00	0.00	0.17	0.51	8.08	8.09	0.09	0.04	0.03
C10	0.08	0.06	0.13	0.08	0.05	0.21	-0.04	0.03	-0.08	-0.08	-0.04	-0.08	-0.02	-0.01	-0.01	8.01	-0.03	-0.08	-0.06	-0.08	-0.05
C11	-0.08	-0.08	0.02	-0.08	-0.08	0.06	8.02	0.01	8.08	0.01	-0.01	-0.03	0.12	0.00	8.01	8.00	-0.01	0.00	0.00	0.00	0.01
C12	0.00	-0.01	0.02	0.00	-0.01	-0.03	0.08	-0.03	-0.01	0.01	0.10	8.21	-0.02	0.02	0.01	8.01	0.02	0.08	8.08	8.15	0.13
M1	0.01	0.02	0.04	-0.01	0.02	0.01	0.08	0.07	-0.05	-0.14	-0.08	8.08	-0.03	0.02	-0.08	-0.03	-0.01	-0.02	-0.02	0.02	0.08
M2	-0.05	-0.05	-0.08	-0.05	-0.05	-0.10	0.03	-0.08	-0.04	8.00	0.17	-0.04	-0.02	0.01	0.01	0.02	-0.01	8.02	0.02	0.14	0.13
M3	0.05	0.04	0.08	-0.15	-0.08	8.08	-0.01	0.01	-0.23	-0.02	9.00	0.02	-0.08	8.08	8.04	0.00	-0.04	0.01	0.01	0.16	0.18
M4	0.56	0.59	0.32	0.55	0.59	0.27	-0.04	8.03	0.07	-0.01	-0.04	-0.13	-0.04	0.00	0.04	-0.05	-0.02	0.02	8.02	0.08	0.08
M5	-0.04	-0.04	-0.09	-0.04	-0.04	0.02	8.00	0.00	8.01	-0.08	0.00	0.00	9.00	0.27	8.01	0.00	-0.08	-0.40	-0.40	8.08	0.03
M6	8.02	0.00	0.02	0.02	0.00	0.03	0.01	-0.01	0.02	0.03	0.05	8.00	0.05	0.08	0.08	0.02	-0.01	0.00	8.00	0.02	0.07
G1	-0.08	-0.08	-0.06	-0.08	-0.08	-0.01	-0.02	-0.19	-0.12	0.18	-0.05	-0.17	0.09	-0.03	0.10	0.11	-0.69	0.09	8.09	0.32	0.33
G2	0.02	0.02	0.10	-0.02	0.00	0.27	-0.07	-0.06	-0.09	8.05	-0.11	-0.13	0.04	-0.01	0.04	0.07	-0.06	0.09	8.09	0.46	0.46
G3	0.09	0.06	0.09	0.06	0.06	0.13	-0.01	-0.17	0.07	9.02	0.02	-0.05	0.01	8.01	0.05	0.81	-0.38	8.00	0.00	0.18	0.11
G4	0.03	0.04	0.05	0.03	0.04	0.03	-0.07	0.06	0.08	-0.01	-0.02	-0.17	-0.08	-0.02	-0.01	-0.02	8.00	-0.03	-0.03	-0.16	-0.16
G5	0.02	0.01	0.03	0.02	0.01	0.02	0.00	0.05	0.02	0.04	0.01	-0.01	0.01	-0.75	8.03	0.01	0.05	1.08	1.00	0.02	8.02
G6	0.03	0.04	-0.02	0.02	0.04	0.01	-0.03	-0.02	-0.05	0.01	9.01	-0.06	-0.01	0.00	0.02	0.01	-0.02	-0.01	-0.01	0.01	8.00
N1	0.10	0.07	0.13	0.06	0.05	0.11	0.00	8.05	8.14	0.10	0.09	-0.04	0.04	0.01	8.12	0.08	-0.04	0.00	8.00	-0.18	-0.11
N2	8.07	0.07	0.11	0.02	0.04	0.02	0.01	0.04	0.04	-0.02	-0.09	-0.04	-0.08	0.07	0.04	-0.03	0.01	-0.08	-0.08	0.02	0.04
N3	8.03	0.04	0.05	0.04	0.05	0.00	0.05	0.13	0.16	-0.04	-0.08	0.08	-0.02	0.07	-0.05	-0.05	0.08	-0.02	-0.02	-0.08	-0.03

	S1	S2	S3	S4	S5	S6	S7	S8	S9	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
S1	1.00																				
S2	8.08	1.09																			
S3	0.08	0.28	1.00																		
S4	0.03	1.00	0.27	1.00																	
S5	0.87	8.01	0.01	0.01	1.00																
S6	8.02	0.89	0.32	0.69	0.00	1.00															
S7	8.01	0.87	0.32	0.87	0.81	0.98	1.00														
S8	0.01	0.83	0.26	0.83	0.02	0.93	8.94	1.00													
S9	0.02	0.90	0.34	0.90	8.00	0.97	0.05	8.91	1.00												
C1	8.01	-0.02	-0.02	-0.02	8.04	-0.03	0.00	-0.01	-0.02	1.00											
C2	8.03	-0.06	-0.03	-0.06	0.10	-0.04	-0.03	-0.04	-0.04	8.22	1.00										
C3	8.00	0.01	0.13	0.01	0.01	0.02	8.02	0.00	8.02	0.02	0.81	1.00									
C4	-0.01	-0.06	-0.19	-0.06	0.00	-0.05	-0.05	-0.02	-0.05	-0.05	-0.01	-0.04	1.00								
C5	-0.01	0.01	0.18	0.01	8.01	0.03	0.03	0.00	0.03	8.04	0.02	8.02	-0.01	1.09							
C6	-0.02	0.00	-0.02	0.00	-0.01	-0.01	-0.01	0.00	-0.01	-0.02	0.00	8.01	0.02	0.01	1.00						
C7	-0.02	0.01	-0.03	0.01	-0.02	-0.03	-0.01	-0.01	-0.03	-0.03	0.00	0.01	0.01	0.01	0.91	1.00					
C8	-0.01	-0.11	-0.26	-0.11	0.00	-0.11	-0.11	-0.09	-0.11	-0.02	0.02	-0.08	0.15	-0.04	-0.14	-0.13	1.00				
C9	-0.01	-0.02	-0.09	-0.01	0.00	-0.05	-0.05	-0.04	-0.05	-0.01	0.00	8.02	0.05	0.01	8.33	8.30	8.34	1.00			
C10	-0.02	0.00	-0.09	0.00	8.02	0.08	-0.01	0.02	0.00	0.04	0.02	-0.01	8.04	-0.02	-0.12	-0.10	8.31	0.16	1.00		
C11	8.08	0.01	-0.01	0.01	0.00	0.08	0.00	0.00	0.00	8.00	0.00	0.00	0.01	-0.01	-0.02	0.00	-0.10	-0.11	8.17	1.00	
C12	0.09	-0.03	-0.69	-0.03	0.01	-0.03	-0.04	-0.08	-0.08	8.00	0.02	8.08	0.08	-0.01	0.08	-0.09	-0.01	0.02	-0.04	-0.81	1.00
M1	8.04	-0.01	-0.02	0.00	0.02	-0.02	-0.04	-0.02	-0.01	-0.02	-0.08	-0.53	-0.04	-0.52	-0.08	-0.07	0.01	0.01	0.08	-0.05	-0.01
M2	-0.01	-0.05	-0.04	-0.05	0.01	-0.07	-0.07	-0.07	-0.08	0.08	0.01	0.09	0.09	0.02	8.07	8.02	8.03	0.01	-0.87	-0.02	0.11
M3	-0.01	0.01	0.00	8.01	0.00	-0.19	-0.18	-0.14	8.02	0.00	0.01	-0.01	0.08	-0.01	9.08	8.02	0.02	0.01	8.01	0.00	0.02
M4	-0.02	-0.03	-0.25	-0.03	0.00	-0.14	-0.13	-0.08	-0.13	-0.04	-0.01	-0.01	0.13	-0.08	-0.02	0.05	0.36	0.54	0.37	0.05	-0.01
M5	0.29	-0.07	8.01	-0.07	0.37	-0.05	-0.04	0.08	-0.04	0.00	0.02	0.00	-0.01	8.09	0.03	0.01	0.01	0.08	8.11	0.00	0.02
M6	0.05	-0.02	-0.08	-0.02	8.10	-0.06	-0.05	-0.04	-0.05	0.07	0.14	8.02	-0.01	0.04	0.04	0.03	0.04	0.00	-0.03	-0.02	0.01
G1	-0.06	-0.24	-0.08	-0.26	-0.05	-0.30	-0.29	-0.25	-0.31	0.05	0.08	8.09	-0.08	0.01	0.24	0.24	-0.15	0.03	-0.21	8.04	0.16
G2	-0.08	-0.09	-0.29	-0.08	-0.02	-0.20	-0.19	-0.09	-0.17	-0.01	0.00	-0.03	8.03	-0.05	8.23	0.21	-0.02	0.69	0.06	0.07	0.20
G3	0.01	-0.73	-0.14	-0.73	0.00	-0.77	-0.78	-0.74	-0.77	-0.06	-0.04	8.00	0.04	0.00	8.14	0.13	8.05	0.05	-0.06	0.00	0.04
G4	0.01	0.05	0.12	0.05	-0.01	0.04	0.05	0.08	0.04	0.00	8.00	8.07	-0.08	0.01	-0.08	0.03	8.02	-0.03	0.07	0.02	-0.40
G5	-0.88	-0.03	-0.11	-0.03	-0.78	-0.02	-0.02	-0.04	-0.02	0.01	-0.01	0.01	8.02	0.01	0.04	0.04	8.02	0.00	-0.08	0.00	0.00
G6	0.00	-0.03	-0.03	-0.03	0.01	-0.01	-0.01	0.14	-0.01	-0.01	-0.01	0.00	0.03	0.00	0.08	0.03	8.02	0.11	0.02	0.01	0.00
N1	0.02	-0.05	-0.03	-0.05	0.01	-0.11	-0.11	-0.13	-0.09	-0.05	-0.02	-0.02	-0.05	-0.04	8.06	0.08	0.01	0.01	0.00	0.06	-0.10
N2	8.04	-0.12	0.02	-0.12	8.05	-0.16	-0.17	-0.16	-0.14	8.05	0.01	0.02	-0.01	0.04	-0.01	-0.02	0.08	-0.04	-0.14	-0.01	-0.01
N3	8.04	0.04	0.04	0.04	0.04	0.08	0.07	0.04	0.04	0.05	0.00	0.04	0.00	0.08	-0.09	-0.09	8.02	-0.10	-0.18	-0.01	-0.03

	M1	M2	M3	M4	M5	M6	G1	G2	G3	G4	G5	G6	N1	N2	N3
M1	1.00														
M2	-0.11	1.00													
M3	8.04	0.04	1.00												
M4	0.02	-0.08	0.04	1.00											
M5	-0.03	0.01	0.00	-0.01	1.90										
M6	-0.03	0.00	8.03	0.01	0.01	1.00									
G1	0.04	0.20	8.03	-0.69	0.04	0.07	1.00								
G2	8.09	0.00	0.20	0.24	0.03	0.08	8.73	1.00							
G3	0.00	0.06	0.05	0.03	0.04	0.04	8.36	0.20	1.00						
G4	0.08	-0.27	-0.04	8.06	-0.01	-0.01	-0.12	-0.10	-0.08	1.00					
G5	-0.02	0.01	0.90	8.02	-0.40	8.00	0.09	0.09	0.00	-0.01	1.00				
G6	0.08	-0.02	0.02	0.13	8.05	0.02	8.01	0.12	0.03	8.02	-0.81	1.00			
N1	-0.06	-0.12	0.07	0.01	-0.04	-0.05	-0.29	-0.21	0.11	0.09	0.01	-0.03	1.00		
N2	-0.05	-0.06	0.18	-0.12	8.05	0.01	-0.01	-0.18	0.10	0.06	-0.03	-0.08	0.11	1.08	
N3	-0.07	-0.06	-0.06	-0.13	-0.03	0.81	-0.16	-0.26	-0.10	0.18	-0.02	-0.11	0.05	0.84	1.00

Appendix G.2. A t-test of all the correlations in Appendix G.1.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	E1	E2	E3	P1	P2	P3	P4	P5	P6	P7	S1	S2	S3	S4	S5	S6	S7	S8	S9	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	M1	M2	M3	M4	M5	M6	G1	G2	G3	G4	G5	G6	N1	N2
L1																																																								
L2																																																								
L3																																																								
L4																																																								
L5																																																								
L6																																																								
L7	-0.7	-0.9	[illegible]	-0.7	-0.9	-0.9																																																		
L8	-1.1	0.9	-0.4	-1.2	-0.1		-0.4																																																	
L9							1.7	0.6																																																
L10	1.8	1.1		1.7	1.2	0.4	0.6	-0.9	1.3																																															
L11	0.2	-0.4	-0.5	0.2	-0.4	0.7	0.1	-0.5	-1.3	0.9																																														
E1	[illegible]	[illegible]	-1.8	[illegible]	[illegible]	0.9		0.3		0.7	0.5																																													
E2	-0.5	-0.3	-1.4	-0.2	-0.1	-1.9		-0.3	1.9	0.2	-0.2	-0.2																																												
E3	-0.1	-0.1	-0.2	-0.1	-0.1	0.0	0.1	-0.1	0.2	-0.1	0.0	0.4	-0.1																																											
P1	[illegible]	[illegible]		[illegible]	[illegible]	1.1	0.1	-0.1	0.6		0.1	0.1	0.3	0.0																																										
P2	0.6	0.4	0.6	0.8	0.4	0.7	0.4	-0.1	0.6		0.1	1.6	0.1	0.0																																										
P3	-0.9	-0.6	-1.5	-0.9	-0.8	-0.6		1.0	-0.8	0.0	-0.3	1.8	-0.3	-0.7	-1.9	0.7																																								
P4	0.3	0.2	0.5	0.3	0.2	0.3	0.2	0.9	0.3	0.6	0.1	0.2	0.2		0.5	0.3	1.6																																							
P5	0.3	0.2	0.5	0.3	0.2	0.3	0.2	0.9	0.3	0.8	0.1	0.2	0.2		0.5	0.3	1.6																																							
P6	0.9	0.6	[illegible]	0.4	0.5		-0.4	-0.7	-0.6	0.3	-1.1	-0.6	-0.6	0.2	0.5	0.4	-0.7	0.4	0.4																																					
P7	0.6	0.8	[illegible]	0.1	0.2	1.9	-0.8	-0.7	-0.8	0.4	-1.6	-1.3	-0.7	0.2	0.6	0.4	-0.8	0.4	0.4																																					
S1	-0.3	-0.2	-0.5	-0.3	-0.2	-0.3	0.0	0.6	-0.1	-0.1	-0.1	0.2	-0.1		-0.1	0.0	-0.4			-0.1	-0.2																																			
S2	-1.1	-1.6	-1.5	-1.2	-1.1	-1.0	-0.1	0.3	-1.0	-0.3	-0.6	-0.8	-0.1	0.4		-0.6		-0.6	-0.6	-1.5	-1.3	0.5																																		
S3							0.2	-1.6	-1.1	0.8	-0.6	-1.1	-0.7	0.4	-0.7	0.7	[illegible]	[illegible]	[illegible]			1.5																																		
S4	-1.1	-1.0	-1.4	-1.2	-1.0	-0.9	-0.1	0.3	-1.0	-0.3	-0.7	-0.6	-0.1	0.4		-0.6		-0.6	-0.6	-1.4	-1.3	0.5																																		
S5	-0.2	-0.1	-0.3	-0.2	-0.1	0.1	0.0	0.6	-0.1	-0.4	0.0	-0.1	-0.1		-0.2	0.6	-0.6			0.3	0.3		0.3	0.3	0.3																															
S6						[illegible]	0.8	1.4	-1.7	-0.5	-0.4	[illegible]	-0.7	0.0		-0.1		-0.4	-0.4			0.3				0.6																														
S7						-1.9	0.8	1.8	-1.2	-0.5	-1.9	[illegible]	-0.7	0.1		-0.1		-0.4	-0.4			0.3				0.1																														
S8						-1.5	1.6	1.6	-1.0	-0.8	-0.8	1.9	-1.9	0.6		-0.1		-0.7	-0.7			0.1				0.3																														
S9						-1.9	1.0	1.6		-0.8	0.0	[illegible]	-1.6	0.6		-0.2		-0.4	-0.4			0.3				0.1																														
C1	-0.6	-0.6	-1.1	-0.7	-0.8	-1.4	-0.2	0.3	-0.1	0.1	-0.3	-0.8	0.3	0.8	0.1	0.1	-0.4	0.2	0.2	0.1	0.3	0.2	-0.3	-0.3	-0.3	0.6	-0.5	0.0	-0.1	-0.4																										
C2	-0.2	-0.2	-0.3	-0.2	-0.2	-0.4	-1.9	-0.1	0.0	0.0	1.3	-0.2	0.3	[illegible]	0.2	0.1	-1.5	-0.2	-0.2	0.4	0.4	0.6	-1.2	-0.5	-1.2	1.7	-0.7	-0.5	-0.7	-0.7																										
C3	-0.1	-0.1	-0.2	-0.1	-0.2	-0.5	-0.3	-0.2	0.6	0.2	0.6	0.3	0.7	0.3	0.0	0.0		0.1	0.1	-1.9	-0.9	-0.1	0.1	[illegible]	0.1	0.2	0.4	0.5	0.0	0.4	0.3	0.2																								
C4	1.1	1.2	0.6	1.6	1.2		0.8	0.5	-0.1	-0.5	0.3	1.2	-0.1	0.0	-0.4	-0.7	0.1	0.4	0.4	1.5	1.4	-0.1	-1.1		-1.1	-0.1	-0.9	-0.9	-0.3	-0.8	-0.9	-0.1	-0.7																							
C5	0.0	-0.3	-0.4	-0.1	-0.3	-0.4	-0.4	-0.7	0.7	0.9	0.4	0.1	0.6	0.3	0.2	0.1		0.2	0.2	-1.7	-1.3	-0.1	0.1		0.1	0.2	0.5	0.5	-0.1	0.6	0.7	0.4		-0.2																						
C6	-1.3	-1.5	[illegible]	-1.3	-1.8	[illegible]	-0.1	-1.2	-0.6	0.6	0.1	-0.8	-0.5	-0.1			0.4	0.8	0.6	[illegible]		-0.4	0.0	-0.4	0.0	-0.3	-0.3	-0.1	0.0	-0.3	-0.4	0.1	0.3	0.3	0.2																					
C7	[illegible]	[illegible]	-0.6	[illegible]	[illegible]		-1.1	-1.6	-0.8	0.7	0.1		-0.5	-0.3			0.0	0.6	0.6	1.8		-0.4	0.1	-0.6	0.1	-0.3	-0.5	-0.3	-0.1	-0.5	-0.5	0.0	0.2	0.2	0.2																					
C8							0.2	1.9	0.0	-0.6	-0.8	0.6	0.5	-0.1	-1.2	-0.3	0.4	0.5	0.5	-0.3	-0.3	-0.2	-1.9		-1.9	-0.1	-1.9	[illegible]	-1.7	-1.9	-0.4	0.4	-1.0		-0.6		[illegible]																			
C9			1.4			1.4	0.2	-0.6	0.6	-0.7	-0.3	-0.2	0.1	0.6			0.5	0.1	0.1	0.7	0.5	-0.1	-0.3	-1.5	-0.3	0.0	-0.9	-1.0	-0.7	-0.9	-0.2	0.1	0.3	1.0	0.2																					
C10	0.6	1.0	[illegible]	0.5	1.0		-0.7	0.6	-0.5	-1.1	-0.7	-1.1	-0.4	-0.2	-0.1	0.1	-0.5	-1.2	-1.2	-0.5	-0.8	-0.4	0.0	-1.6	0.6	0.3	0.0	-0.2	0.3	0.1	0.8	0.4	-0.1	0.7	-0.4	[illegible]	-1.9																			
C11	-1.5	-1.4	0.3	-1.5	-1.4	1.1	0.4	0.1	0.5	0.3	-0.1	-0.5	[illegible]	0.6	0.1	0.6	-0.2	0.0	0.6	-0.1	0.2	-0.1	0.2	-0.2	0.2	0.0	-0.1	0.0	0.0	0.0	0.9	0.0	0.6	0.1	-0.1	-0.3	0.0	-1.9																		
C12	0.1	-0.1	0.3	0.6	-0.2	-0.5	1.0	-0.5	-0.2	0.2	1.8		-0.4	0.4	0.2	0.3	0.3	0.5	0.5		[illegible]	-0.1	-0.6	-1.6	-0.5	0.2	-0.6	-0.7	-0.5	-0.5	0.6	0.3	0.0	1.5	-0.1	1.1	-1.6	-0.2	0.4	-0.8	-0.2															
M1	0.1	0.4	0.7	-0.2	0.3	0.1	0.5	1.3	-0.6		-1.4	0.5	-0.5	0.4	-1.1	-0.5	-0.2	-0.3	-0.3	0.3	0.1	0.7	-0.1	-0.4	-0.1	0.4	-0.5	-0.7	-0.3	-0.2	-0.4	-1.1		-0.7		-1.4	-1.2	0.1	0.2	1.4	-1.0	-0.1														
M2	-0.8	-0.8	-1.4	-1.0	-0.9	-1.7	0.6	-1.0	-0.7	0.1		-0.6	-0.4	0.1	0.2	0.3	-0.2	0.4	0.4	[illegible]	[illegible]	-0.1	-0.9	-0.8	-0.6	0.2	-1.2	-1.3	-1.2	-1.1	0.1	0.2	0.1	1.8	0.3	1.4	0.3	0.6	0.1	-1.3	-0.3	[illegible]	[illegible]													
M3	0.6	0.7	1.4		-1.5	1.1	-0.2	0.2		-0.3	0.1	0.4	-1.4	0.0	0.6	0.1	-0.7	0.1	0.1			-0.2	0.1	0.0	0.2	0.0			[illegible]	0.4	0.1	0.2	-0.2	0.8	-0.2	0.5	0.4	0.4	0.3	0.2	0.0	0.4	0.7	0.8												
M4							-0.7	0.6	1.2	-0.2	-0.6	[illegible]	-0.7	-0.1	0.7	-1.6	-0.3	0.4	0.4	1.4	1.2	-0.3	-0.6		-0.5	-0.1	[illegible]	[illegible]	-1.5	[illegible]	-0.7	-0.2	-0.2	[illegible]	-0.6	-0.4	1.0			1.0	-0.2	0.4	-1.5	0.7												
M5	-0.7	-0.7	-1.6	-0.7	-0.7	0.4	0.1	0.0	0.1	-1.4	-0.1	0.0	-0.1		0.1	0.0	-0.8			0.6	0.6		-1.3	0.1	-1.3		-0.9	-0.6	0.5	-0.6	0.9	0.4	0.1	-0.3	0.1	0.5	0.2	0.1	1.1	[illegible]	0.1	0.4	-1.0	0.1	0.1	-0.2										
M6	0.4	0.6	0.3	0.3	0.0	0.6	0.1	-0.1	0.4	0.5	0.9	0.0	1.9	1.5	0.5	0.3	-0.3	0.0	0.0	0.4	1.3	0.9	-0.3	-1.4	-0.3	1.7	-1.8	-0.8	-0.7	-0.8	1.2	[illegible]	0.4	-0.2	0.6	0.7	0.6	0.7	0.0	-0.5	-0.3	0.2	-0.6	0.0	0.5	0.2	0.2									
G1	-1.1	-1.4	-1.1	-1.2	-1.4	-0.3	-0.4		[illegible]	1.9	-1.0		1.6	-0.6	1.6	[illegible]	-1.6	1.5	1.5			-1.0		-1.4		-0.9					0.9	0.5	0.0	-1.1	0.2				0.6		0.8		0.7		0.5	-1.6	0.8	1.2								
G2	0.3	0.4	1.6	-0.4	-0.1		-1.3	-1.0	-1.6	1.0		[illegible]	0.7	-0.2	0.7	1.3	-1.1	1.6	1.8			-1.1	-1.6		-1.4	-0.4			-1.7		-0.3	0.1	-0.6	0.8	-1.6			-0.3	1.5	1.1	1.3		1.7	0.1			0.5	1.4								
G3	1.6	1.1	1.7	1.4	1.0	[illegible]	-0.1		1.3	0.4	0.3	-0.9	0.2	0.2	0.9	0.3		0.1	0.1	1.8	[illegible]	0.1				0.1					-1.1	-0.7	0.6	0.8	0.8	[illegible]	[illegible]	0.9	0.9	-1.1	0.0	0.7	0.1	1.0	0.9	0.6	0.7	0.7								
G4	0.5	0.8	1.0	0.6	0.6	0.5	-1.3	1.0	0.5	-0.2	-0.3		-1.4	-0.4	-0.2	-0.4	0.1	-0.5	-0.5			0.1	1.0	[illegible]	1.0	-0.1	0.8	0.9	1.5	0.7	0.0	0.6	1.3	-1.1	0.2	-1.4	0.5	0.3	-0.5	1.3	0.3		1.1		-0.7	1.1	-0.2	-0.2	[illegible]	-1.9	-1.1					
G5	0.3	0.2	0.5	0.4	0.3	0.3	0.1	0.9	0.3	0.8	0.1	-0.1	0.2		0.5	0.3	0.9			0.3	0.4		-0.6	[illegible]	-0.8		-0.4	-0.4	-0.7	-0.4	0.2	-0.2	0.1	0.3	0.2	0.7	0.6	0.4	0.0	-1.2	0.1	0.0	-0.3	0.2	0.1	0.4		0.8	1.6	1.6	0.0	-0.1				
G6	0.5	0.7	-0.3	0.4	0.7	0.2	-0.8	-0.3	-0.9	0.2	0.2	-1.1	-0.2	0.0	0.3	0.2	-0.4	-0.1	-0.1	0.2	0.0	-0.1	-0.9	-0.8	-0.6	0.2	-0.3	-0.1	[illegible]	-0.2	-0.2	-0.1	0.0	0.6	0.9	0.8	0.5	0.3	[illegible]	0.3	0.2	-0.1	1.4	-0.3	0.3	[illegible]	1.6	0.4	0.3	[illegible]	0.5	0.4	-0.2			
N1	1.6	1.2	[illegible]	1.5	1.6	[illegible]	0.0	0.9	[illegible]	1.6	1.8	-0.6	0.7	0.2	[illegible]	1.5	-0.7	0.1	0.1	-1.8	[illegible]	0.3	-1.0	-0.5	-0.9	0.2		-1.9	[illegible]	-1.7	-1.0	-0.4	-0.3	-0.9	-0.6	1.0	1.5	0.1	0.2	-0.1	1.5	-1.8	-1.2	[illegible]	1.2	0.2	-0.8	-0.9				1.7	0.2	-0.5		
N2	1.3	1.2	1.9	0.4	0.7	0.3	0.1	0.6	0.8	-0.4	-1.6	-0.7	-0.8	1.2	0.7	-0.6	0.1	-0.6	-0.6	0.3	0.6	0.9	[illegible]		0.4	0.9					0.9	0.1	0.3	-0.1	0.7	-0.2	-0.4	0.5	-0.6	[illegible]	-0.1	-0.2	-0.8	-1.0	[illegible]		0.9	0.2	-0.2	-1.6	1.8	1.0	-0.6	-1.5	[illegible]	
N3	0.6	0.7	0.9	0.8	0.9	-0.1	0.8	[illegible]		-0.7	-0.5	1.4	-0.3	1.2	-0.9	-0.9	1.4	-0.4	-0.4	-1.1	-0.6	0.7		0.7	0.8	0.8	0.6	1.1	1.3	0.8	0.8	0.9	0.0	0.7	0.1	1.6	-1.5	-1.8	0.4	-1.9		-0.2	-0.8	-1.3	-1.1	-1.6	[illegible]	-0.5	0.2			-1.6	1.7	-0.3	[illegible]	1.0

APPENDIX H – FEATURE SUBSETS SELECTED

Appendix H presents the ten different binary masks identified using PBIL for each of the three forecast periods as follows:

- **Appendix H.1**: 1-Yr Fwd feature subsets
- **Appendix H.2**: 2-Yr Fwd feature subsets
- **Appendix H.3**: 3-Yr Fwd feature subsets

The highlighted binary masks are the optimal feature subsets used as inputs into the classification models constructed in the study. They are referred in the report as follows:

- **1-Yr Fwd Model**: Optimal Feature Subset A = Appendix H.1. mask A1
- **2-Yr Fwd Model**: Optimal Feature Subset A = Appendix H.2. mask B1; Optimal Feature Subset B = Appendix H.2. mask B5
- **3-Yr Fwd Model**: Optimal Feature Subset A = Appendix H.3. mask C5; Optimal Feature Subset B = Appendix H.3. mask C6; Optimal Feature Subset C = Appendix H.3. mask C10

Finally, Appendix H.4.

Appendix H.1. Binary masks for 1 Year Forward Forecasting Models

		A2	A3	A4	A5	A6	A7	A8	A9	A10	
L1(1)	L1	0	1	1	1	0	1	0	0	0	Current Ratio
L2(1)	L2	1	0	0	1	0	0	0	0	0	Quick Ratio
L3(1)	L3	0	1	1	1	1	0	1	1	0	Cash Ratio
L4(1)	L4	0	0	0	0	1	1	1	1	0	Non-Group Current Ratio
L5(1)	L5	1	1	1	1	1	0	1	1	0	Non-Group Quick Ratio
L6(1)	L6	0	1	1	0	1	1	1	0	0	Cash to Total Assets Ratio
L7(1)	L7	0	0	0	0	0	0	1	0	0	Receivables Turnover
L8(1)	L8	1	1	1	1	1	1	1	0	1	Inventory Turnover
L9(1)	L9	0	1	1	0	0	0	0	1	0	Payables Turnover
L10(1)	L10	1	1	1	0	1	0	1	0	0	Cash Conversion Cycle
L11(1)	L11	0	0	0	1	0	1	1	1	0	Preference Dividend Declared?
E1(1)	E1	1	0	0	0	0	0	0	1	0	Total Asset Turnover
E2(1)	E2	0	0	0	0	1	1	1	0	0	Fixed Asset Turnover
E3(1)	E3	1	1	1	1	0	0	1	1	0	Equity Turnover
P1(1)	P1	1	0	0	0	1	1	0	0	0	Operating Profit Margin
P2(1)	P2	1	1	1	0	0	0	1	1	1	Net Profit Margin
P3(1)	P3	0	0	0	0	0	0	0	0	0	Return on Total Capital
P4(1)	P4	0	0	0	0	1	0	0	0	1	Return on Total Equity
P5(1)	P5	1	0	0	0	0	1	0	1	1	Return on Ordinary Equity
P6(1)	P6	0	0	0	0	0	1	0	1	0	Basic EPS
P7(1)	P7	0	0	0	0	0	0	0	0	0	Headline EPS
S1(1)	S1	0	0	0	1	1	0	0	0	0	Debt Equity Ratio
S2(1)	S2	1	0	0	0	1	1	1	1	1	Long Term Debt to Assets Ratio
S3(1)	S3	0	0	0	0	0	1	0	0	0	Interest Bearing Debt to Asset Ratio
S4(1)	S4	1	1	1	1	0	1	1	1	0	External Long Term Debt to Asset Ratio
S5(1)	S5	0	0	0	0	0	0	0	0	0	Financial Leverage
S6(1)	S6	1	1	1	1	1	1	1	1	1	Total Debt to Assets Ratio
S7(1)	S7	0	0	0	0	0	0	1	1	1	Total Debt and Contingencies to Assets Ratio
S8(1)	S8	0	0	0	0	0	1	0	0	0	Total Commitments to Assets Ratio
S9(1)	S9	1	1	1	1	1	1	1	0	0	External Debt to Assets Ratio
C1(1)	C1	1	1	1	0	0	0	0	0	0	Interest Coverage Ratio (Inverse)
C2(1)	C2	0	0	0	0	0	0	0	0	0	Fixed Charge Coverage Ratio (Inverse)
C3(1)	C3	1	1	1	0	1	1	1	1	1	Interest Expense to Cash Flow Ratio
C4(1)	C4	1	1	1	0	1	1	0	1	1	Cash Flow to Long Term Debt Ratio
C5(1)	C5	1	0	0	1	0	0	0	1	1	Cash Flow to Interest Bearing Debt (Inverse)
C6(1)	C6	1	0	0	0	0	0	0	0	0	Cash Flow to Total Debt Ratio
C7(1)	C7	1	0	0	1	1	0	0	1	0	Cash Flow from Operations to Total Debt
C8(1)	C8	1	1	1	1	0	0	0	0	1	Change in Long Term Debt to Long Term Debt
C9(1)	C9	0	0	0	0	0	0	0	0	0	Change in Total Debt to Total Debt
C10(1)	C10	1	0	0	0	1	0	1	1	1	Proceeds on Share Issue to Total Assets
C11(1)	C11	1	0	0	1	1	0	0	0	1	Cash Invested in Investing Activities to Fixed Assets
C12(1)	C12	1	0	0	0	1	1	0	1	0	Proportion of Dividend that is Non-Cash
M1(1)	M1	1	0	0	1	1	0	0	1	1	Trading Turnover
M2(1)	M2	0	1	1	1	0	0	1	0	0	Total Dividend Yield
M3(1)	M3	1	0	0	0	1	1	1	1	0	Earnings Yield
M4(1)	M4	0	1	1	0	0	0	0	0	0	Market Value of Equity to Total Debt
M5(1)	M5	1	1	1	0	0	1	1	0	0	Price Book Ratio
M6(1)	M6	0	0	0	0	0	0	0	1	0	Capital to EBITDA Ratio
M7(1)	M7	0	0	0	1	1	0	1	1	0	Excess Return (1 Fin Yr)
M8(1)	M8	0	0	0	1	0	0	0	0	0	Excess Return (2 Fin Yrs)
G1(1)	G1	1	0	0	0	1	0	1	1	1	Log(Assets)
G2(1)	G2	0	1	1	1	1	1	1	1	1	Log(Market Capitalisation)
G3(1)	G3	1	0	0	1	1	0	1	1	1	Retained Income to Total Assets Ratio
G4(1)	G4	0	1	1	0	0	0	1	1	1	Retention Ratio
G5(1)	G5	1	1	1	0	0	1	1	1	0	Growth Rate
G6(1)	G6	0	0	0	0	1	1	0	0	1	Capital Commitments to Fixed Assets Ratio
N1(1)	N1	1	0	0	1	1	1	1	1	0	Directors Shareholding
N2(1)	N2	0	0	0	0	1	1	0	0	0	Auditors Report Qualification
N3(1)	N3	0	0	0	1	0	1	0	0	0	Controlled by Another Company
L1(2)	L1	0	1	1	0	1	0	0	1	1	Current Ratio
L2(2)	L2	1	1	1	1	1	0	1	0	1	Quick Ratio
L3(2)	L3	0	0	0	0	0	0	1	1	0	Cash Ratio
L4(2)	L4	0	1	1	0	1	1	1	1	0	Non-Group Current Ratio
L5(2)	L5	0	1	1	0	1	0	0	0	1	Non-Group Quick Ratio
L6(2)	L6	1	1	1	1	0	1	0	0	1	Cash to Total Assets Ratio
L7(2)	L7	1	1	1	0	0	1	0	1	0	Receivables Turnover
L8(2)	L8	0	0	0	1	0	0	1	1	1	Inventory Turnover
L9(2)	L9	0	0	0	0	1	1	0	1	0	Payables Turnover
L10(2)	L10	0	0	0	0	0	0	1	1	0	Cash Conversion Cycle
L11(2)	L11	0	0	0	0	0	1	1	0	0	Preference Dividend Declared?
E1(2)	E1	1	0	0	1	1	0	1	0	0	Total Asset Turnover
E2(2)	E2	0	0	0	0	0	0	0	0	0	Fixed Asset Turnover
E3(2)	E3	1	1	1	0	0	1	1	0	1	Equity Turnover
P1(2)	P1	0	0	0	0	1	0	0	0	1	Operating Profit Margin
P2(2)	P2	1	0	0	1	1	0	1	1	0	Net Profit Margin
P3(2)	P3	0	0	0	0	0	0	0	0	1	Return on Total Capital
P4(2)	P4	1	1	1	0	1	1	1	1	1	Return on Total Equity
P5(2)	P5	0	1	1	0	1	0	0	1	0	Return on Ordinary Equity
P6(2)	P6	1	1	1	1	1	0	1	0	1	Basic EPS
P7(2)	P7	1	0	0	1	0	1	1	1	1	Headline EPS
S1(2)	S1	0	0	0	0	0	1	1	1	1	Debt Equity Ratio
S2(2)	S2	1	1	1	0	1	1	0	0	0	Long Term Debt to Assets Ratio
S3(2)	S3	1	0	0	0	0	1	0	0	1	Interest Bearing Debt to Asset Ratio
S4(2)	S4	0	1	1	0	1	1	1	0	1	External Long Term Debt to Asset Ratio
S5(2)	S5	0	0	0	1	1	0	0	0	0	Financial Leverage
S6(2)	S6	1	0	0	0	0	1	1	1	0	Total Debt to Assets Ratio
S7(2)	S7	0	0	0	0	0	0	0	1	1	Total Debt and Contingencies to Assets Ratio
S8(2)	S8	0	0	0	0	0	0	0	1	0	Total Commitments to Assets Ratio
S9(2)	S9	1	1	1	1	0	1	0	1	0	External Debt to Assets Ratio
C1(2)	C1	1	1	1	0	1	0	0	0	1	Interest Coverage Ratio (Inverse)
C2(2)	C2	1	1	1	0	1	0	0	0	1	Fixed Charge Coverage Ratio (Inverse)
C3(2)	C3	1	0	0	1	0	1	1	0	1	Interest Expense to Cash Flow Ratio
C4(2)	C4	0	1	1	1	0	1	1	0	0	Cash Flow to Long Term Debt Ratio
C5(2)	C5	0	0	0	0	0	0	1	0	1	Cash Flow to Interest Bearing Debt (Inverse)
C6(2)	C6	1	0	0	1	1	0	1	1	0	Cash Flow to Total Debt Ratio
C7(2)	C7	0	1	1	1	1	1	1	1	1	Cash Flow from Operations to Total Debt
C8(2)	C8	1	0	0	1	0	0	0	1	1	Change in Long Term Debt to Long Term Debt
C9(2)	C9	0	0	0	0	0	1	1	0	1	Change in Total Debt to Total Debt
C10(2)	C10	1	1	1	0	1	1	0	0	1	Proceeds on Share Issue to Total Assets
C11(2)	C11	1	1	1	1	0	0	0	0	0	Cash Invested in Investing Activities to Fixed Assets
C12(2)	C12	0	0	0	0	0	0	0	0	0	Proportion of Dividend that is Non-Cash
M1(2)	M1	1	0	0	0	0	1	1	0	1	Trading Turnover
M2(2)	M2	1	1	1	0	1	1	1	1	1	Total Dividend Yield
M3(2)	M3	0	1	1	0	1	0	0	0	0	Earnings Yield
M4(2)	M4	1	1	1	1	0	0	1	1	0	Market Value of Equity to Total Debt
M5(2)	M5	0	0	0	0	1	1	0	1	0	Price Book Ratio
M6(2)	M6	1	1	1	1	1	1	1	1	1	Capital to EBITDA Ratio
M7(2)	M7	0	0	0	1	0	1	0	1	0	Excess Return (1 Fin Yr)
G1(2)	G1	1	0	0	1	1	1	1	0	1	Log(Assets)
G2(2)	G2	0	1	1	0	1	1	0	1	1	Log(Market Capitalisation)
G3(2)	G3	1	1	1	1	0	1	1	1	0	Retained Income to Total Assets Ratio
G4(2)	G4	1	0	0	0	0	1	1	1	1	Retention Ratio
G5(2)	G5	0	1	1	1	0	0	1	1	0	Growth Rate
G6(2)	G6	0	0	0	0	0	0	1	0	1	Capital Commitments to Fixed Assets Ratio
N1(2)	N1	1	1	1	0	0	0	0	0	1	Directors Shareholding
N2(2)	N2	0	1	1	1	0	0	0	0	1	Auditors Report Qualification
N3(2)	N3	0	0	0	1	0	0	0	0	1	Controlled by Another Company
L1(3)	L1	0	0	0	1	1	0	1	1	0	Current Ratio
L2(3)	L2	0	1	1	1	1	0	0	0	0	Quick Ratio
L3(3)	L3	0	1	1	0	0	1	0	0	1	Cash Ratio
L4(3)	L4	0	1	1	1	0	0	0	0	0	Non-Group Current Ratio
L5(3)	L5	0	0	0	0	0	1	0	1	0	Non-Group Quick Ratio
L6(3)	L6	0	0	0	0	1	0	1	1	0	Cash to Total Assets Ratio
L7(3)	L7	1	1	1	0	0	0	1	0	0	Receivables Turnover
L8(3)	L8	0	0	0	0	0	0	0	0	0	Inventory Turnover
L9(3)	L9	0	0	0	0	0	1	0	1	1	Payables Turnover
L10(3)	L10	1	0	0	0	0	0	0	1	0	Cash Conversion Cycle
L11(3)	L11	0	1	1	1	1	0	1	1	1	Preference Dividend Declared?
E1(3)	E1	1	0	0	1	0	1	0	1	1	Total Asset Turnover
E2(3)	E2	1	0	0	0	0	0	0	0	0	Fixed Asset Turnover
E3(3)	E3	1	0	0	0	1	1	1	0	0	Equity Turnover
P1(3)	P1	0	1	1	1	1	1	1	0	1	Operating Profit Margin
P2(3)	P2	0	0	0	1	1	1	0	0	1	Net Profit Margin
P3(3)	P3	1	1	1	0	0	1	1	0	1	Return on Total Capital
P4(3)	P4	1	0	0	1	0	0	1	1	0	Return on Total Equity
P5(3)	P5	0	0	0	0	0	0	0	0	1	Return on Ordinary Equity
P6(3)	P6	1	1	1	0	1	1	0	0	1	Basic EPS
P7(3)	P7	0	1	1	0	1	1	1	1	0	Headline EPS
S1(3)	S1	0	1	1	1	0	0	0	0	1	Debt Equity Ratio
S2(3)	S2	1	0	0	1	0	0	0	1	1	Long Term Debt to Assets Ratio
S3(3)	S3	0	0	0	0	1	0	1	1	0	Interest Bearing Debt to Asset Ratio
S4(3)	S4	0	1	1	1	1	0	1	0	0	External Long Term Debt to Asset Ratio
S5(3)	S5	0	0	0	0	0	0	0	0	1	Financial Leverage
S6(3)	S6	0	1	1	0	0	0	1	1	1	Total Debt to Assets Ratio
S7(3)	S7	1	1	1	1	1	1	0	1	1	Total Debt and Contingencies to Assets Ratio
S8(3)	S8	0	0	0	0	0	0	0	0	0	Total Commitments to Assets Ratio
S9(3)	S9	1	0	0	1	1	1	1	1	1	External Debt to Assets Ratio
C1(3)	C1	1	0	0	0	0	1	0	1	1	Interest Coverage Ratio (Inverse)
C2(3)	C2	1	0	0	1	1	1	1	1	1	Fixed Charge Coverage Ratio (Inverse)
C3(3)	C3	1	0	0	0	0	0	0	1	1	Interest Expense to Cash Flow Ratio
C4(3)	C4	1	1	1	0	0	0	1	1	0	Cash Flow to Long Term Debt Ratio
C5(3)	C5	0	0	0	0	0	1	0	1	1	Cash Flow to Interest Bearing Debt (Inverse)
C6(3)	C6	1	1	1	0	1	0	1	1	0	Cash Flow to Total Debt Ratio
C7(3)	C7	1	1	1	1	1	1	1	1	1	Cash Flow from Operations to Total Debt
C8(3)	C8	1	1	1	1	0	1	1	1	1	Change in Long Term Debt to Long Term Debt
C9(3)	C9	0	0	0	0	0	0	0	0	0	Change in Total Debt to Total Debt
C10(3)	C10	1	1	1	0	0	0	0	1	0	Proceeds on Share Issue to Total Assets
C11(3)	C11	0	1	1	0	1	1	1	1	0	Cash Invested in Investing Activities to Fixed Assets
C12(3)	C12	0	0	0	0	0	0	0	1	0	Proportion of Dividend that is Non-Cash
M1(3)	M1	0	0	0	1	0	0	0	0	0	Trading Turnover
M2(3)	M2	1	1	1	1	1	1	1	1	1	Total Dividend Yield
M3(3)	M3	0	0	0	1	0	1	1	0	0	Earnings Yield
M4(3)	M4	0	1	1	1	1	1	1	0	0	Market Value of Equity to Total Debt
M5(3)	M5	0	1	1	1	0	1	0	0	0	Price Book Ratio
M6(3)	M6	1	1	1	0	1	1	1	0	1	Capital to EBITDA Ratio
G1(3)	G1	1	1	1	0	0	1	1	0	0	Log(Assets)
G2(3)	G2	0	0	0	0	0	0	0	1	0	Log(Market Capitalisation)
G3(3)	G3	0	0	0	1	0	0	1	0	0	Retained Income to Total Assets Ratio
G4(3)	G4	1	0	0	1	1	1	0	0	1	Retention Ratio
G5(3)	G5	0	0	0	1	0	0	0	1	0	Growth Rate
G6(3)	G6	1	1	1	1	0	0	0	0	0	Capital Commitments to Fixed Assets Ratio
N1(3)	N1	0	1	1	1	1	1	0	1	1	Directors Shareholding
N2(3)	N2	1	0	0	1	0	0	0	1	1	Auditors Report Qualification
N3(3)	N3	1	1	1	0	0	0	1	0	0	Controlled by Another Company
R1	R1	0	0	0	0	0	0	1	0	1	Business Risk (3yrs)
R2	R2	1	0	0	1	1	0	1	1	1	Sales Variability (3yrs)
R3	R3	0	0	0	0	0	0	1	0	0	Operating Leverage (3yrs)
N4	N4	1	1	1	1	1	1	1	0	1	Number of Years Listed
N5	N5	1	1	1	1	1	0	1	0	0	Number of Years in Existence
SI Values		90%	89%	89%	90%	89%	88%	89%	88%	89%	

Appendix H.1. Binary masks for 1 Year Forward Forecasting Models

Appendix H.2. Binary masks for 2 Year Forward Forecasting Models

		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	
L1(2)	L1	0	0	0	0	0	0	0	0	1	1	Current Ratio
L2(2)	L2	0	0	0	1	0	1	0	0	0	0	Quick Ratio
L3(2)	L3	0	1	1	1	1	0	1	0	1	0	Cash Ratio
L4(2)	L4	0	0	0	1	0	0	0	0	1	0	Non Group Current Ratio
L5(2)	L5	0	0	0	0	0	1	0	0	0	0	Non Group Quick Ratio
L6(2)	L6	0	1	1	1	1	1	0	0	1	1	Cash to Total Assets Ratio
L7(2)	L7	0	0	0	1	1	0	1	1	1	1	Receivables Turnover
L8(2)	L8	0	1	1	1	1	0	1	1	1	0	Inventory Turnover
L9(2)	L9	1	1	1	0	1	0	1	1	1	0	Payables Turnover
L10(2)	L10	1	1	1	1	0	1	1	1	0	1	Cash Conversion Cycle
L11(2)	L11	1	1	1	0	1	1	1	0	0	1	Preference Dividend Declared?
E1(2)	E1	1	1	1	1	0	1	1	1	1	0	Total Asset Turnover
E2(2)	E2	1	0	0	1	1	0	0	1	1	0	Fixed Asset Turnover
E3(2)	E3	0	0	0	1	0	0	0	0	0	1	Equity Turnover
P1(2)	P1	1	1	1	0	0	1	1	0	0	0	Operating Profit Margin
P2(2)	P2	0	0	0	0	1	0	0	1	0	1	Net Profit Margin
P3(2)	P3	0	0	0	1	0	0	0	0	0	0	Return on Total Capital
P4(2)	P4	0	1	1	1	1	1	1	1	1	1	Return on Total Equity
P5(2)	P5	1	0	0	1	1	0	0	1	0	0	Return on Ordinary Equity
P6(2)	P6	1	1	1	0	1	1	1	1	0	1	Basic EPS
P7(2)	P7	0	0	0	0	1	1	0	0	1	1	Headline EPS
S1(2)	S1	0	1	1	1	0	0	0	1	0	1	Debt Equity Ratio
S2(2)	S2	1	1	1	0	0	0	1	1	1	1	Long Term Debt to Assets Ratio
S3(2)	S3	1	0	0	1	0	0	1	1	0	0	Interest Bearing Debt to Asset Ratio
S4(2)	S4	0	1	1	1	1	1	0	1	0	0	External Long Term Debt to Asset Ratio
S5(2)	S5	1	1	1	1	1	1	0	1	0	1	Financial Leverage
S6(2)	S6	0	0	0	1	1	0	1	1	1	1	Total Debt to Assets Ratio
S7(2)	S7	0	0	0	0	1	1	1	1	1	1	Total Debt and Contingencies to Assets Ratio
S8(2)	S8	0	0	0	0	0	0	1	0	1	0	Total Commitments to Assets Ratio
S9(2)	S9	0	1	1	0	1	1	1	1	1	1	External Debt to Assets Ratio
C1(2)	C1	0	0	0	0	0	0	0	1	0	0	Interest Coverage Ratio (Inverse)
C2(2)	C2	0	1	1	0	1	0	1	0	1	1	Fixed Charge Coverage Ratio (Inverse)
C3(2)	C3	1	1	1	1	1	1	1	1	1	1	Interest Expense to Cash Flow Ratio
C4(2)	C4	1	1	1	0	0	0	0	0	0	1	Cash Flow to Long Term Debt Ratio
C5(2)	C5	0	0	0	0	0	0	0	0	0	0	Cash Flow to Interest Bearing Debt (Inverse)
C6(2)	C6	0	1	1	0	0	1	1	0	0	0	Cash Flow to Total Debt Ratio
C7(2)	C7	1	0	0	0	1	1	1	1	0	1	Cash Flow from Operations to Total Debt
C8(2)	C8	1	1	1	0	0	1	0	1	0	0	Change in Long Term Debt to Long Term Debt
C9(2)	C9	1	0	0	1	0	1	1	1	0	0	Change in Total Debt to Total Debt
C10(2)	C10	0	1	1	0	0	1	1	0	1	0	Proceeds on Share Issue to Total Assets
C11(2)	C11	0	1	1	1	1	0	0	0	0	1	Cash Invested in Investing Activities to Fixed Assets
C12(2)	C12	0	0	0	0	0	0	0	0	0	1	Proportion of Dividend that is Non Cash
M1(2)	M1	1	1	1	1	1	0	1	1	1	1	Trading Turnover
M2(2)	M2	0	0	0	0	0	0	0	0	0	0	Total Dividend Yield
M3(2)	M3	0	1	1	0	0	0	0	0	1	0	Earnings Yield
M4(2)	M4	0	1	1	1	0	1	0	0	1	0	Market Value of Equity to Total Debt
M5(2)	M5	0	0	0	1	0	0	1	0	0	0	Price Book Ratio
M6(2)	M6	1	1	1	1	1	0	1	0	1	1	Capital to EBITDA Ratio
M7(2)	M7	1	0	0	0	0	1	0	0	1	1	Excess Return (1 Yr)
G1(2)	G1	0	0	0	1	0	0	0	1	0	0	Log(Assets)
G2(2)	G2	0	0	0		0	0	0	0	0	0	Log(Market Capitalisation)
G3(2)	G3	0	1	1	0	0	0	1	0	0	0	Retained Income to Total Assets Ratio
G4(2)	G4	1	1	1	1	0	1	0	1	0	1	Retention Ratio
G5(2)	G5	1	1	1	0	1	1	1	1	0	0	Growth Rate
G6(2)	G6	0	0	0	0	0	0	0	0	1	0	Capital Commitments to Fixed Assets Ratio
N1(2)	N1	0	0	0	0	0	0	0	1	0	0	Directors Shareholding
N2(2)	N2	0	0	0	1	0	0	0	0	0	1	Auditors Report Qualification
N3(2)	N3	0	1	1	0	1	0	1	0	0	0	Controlled by Another Company
L1(3)	L1	0	0	0	1	0	1	0	0	0	1	Current Ratio
L2(3)	L2	0	1	1	1	0	1	0	0	0	0	Quick Ratio
L3(3)	L3	1	0	0	0	0	0	1	0	1	1	Cash Ratio
L4(3)	L4	0	0	0	1	0	1	0	0	0	0	Non Group Current Ratio
L5(3)	L5	0	0	0	0	1	0	0	0	0	0	Non Group Quick Ratio
L6(3)	L6	0	1	1	1	0	0	0	0	0	0	Cash to Total Assets Ratio
L7(3)	L7	0	0	0	0	0	0	0	0	0	0	Receivables Turnover
L8(3)	L8	0	0	0	0	0	0	1	0	0	1	Inventory Turnover
L9(3)	L9	1	0	0	1	1	1	1	0	1	0	Payables Turnover
L10(3)	L10	0	0	0	0	0	0	0	1	1	1	Cash Conversion Cycle
L11(3)	L11	0	0	0	0	0	0	1	0	1	1	Preference Dividend Declared?
E1(3)	E1	1	0	0	1	1	0	0	1	1	0	Total Asset Turnover
E2(3)	E2	0	0	0	0	1	0	0	0	1	0	Fixed Asset Turnover
E3(3)	E3	0	1	1	1	0	0	0	0	0	0	Equity Turnover
P1(3)	P1	0	1	1	1	1	1	0	0	1	1	Operating Profit Margin
P2(3)	P2	1	1	1	1	0	1	0	0	0	0	Net Profit Margin
P3(3)	P3	0	0	0	1	0	1	1	1	1	1	Return on Total Capital
P4(3)	P4	1	0	0	1	1	1	1	1	1	0	Return on Total Equity
P5(3)	P5	1	1	1	0	0	0	0	0	1	1	Return on Ordinary Equity
P6(3)	P6	0	0	0	0	1	1	0	0	0	0	Basic EPS
P7(3)	P7	0	1	1	0	0	0	1	1	1	0	Headline EPS
S1(3)	S1	1	0	0	0	1	0	1	0	0	1	Debt Equity Ratio
S2(3)	S2	1	1	1	1	1	0	1	1	1	0	Long Term Debt to Assets Ratio
S3(3)	S3	1	0	0	1	1	0	0	1	1	1	Interest Bearing Debt to Asset Ratio
S4(3)	S4	0	0	0	1	1	0	1	1	0	1	External Long Term Debt to Asset Ratio
S5(3)	S5	0	0	0	1	0	0	1	0	1	1	Financial Leverage
S6(3)	S6	0	1	1	0	0	0	0	1	1	1	Total Debt to Assets Ratio
S7(3)	S7	1	1	1	0	1	1	0	1	1	0	Total Debt and Contingencies to Assets Ratio
S8(3)	S8	0	0	0	1	0	1	0	0	0	0	Total Commitments to Assets Ratio
S9(3)	S9	0	1	1	0	0	0	1	0	0	0	External Debt to Assets Ratio
C1(3)	C1	0	1	1	1	0	1	1	0	0	1	Interest Coverage Ratio (Inverse)
C2(3)	C2	0	1	1	1	0	0	0	0	0	1	Fixed Charge Coverage Ratio (Inverse)
C3(3)	C3	0	0	0	0	0	0	0	0	0	0	Interest Expense to Cash Flow Ratio
C4(3)	C4	0	0	0	1	0	0	0	0	0	1	Cash Flow to Long Term Debt Ratio
C5(3)	C5	0	0	0	0	0	0	0	0	0	0	Cash Flow to Interest Bearing Debt (Inverse)
C6(3)	C6	0	1	1	0	0	0	1	1	1	1	Cash Flow to Total Debt Ratio
C7(3)	C7	1	1	1	1	1	0	1	1	1	0	Cash Flow from Operations to Total Debt
C8(3)	C8	1	1	1	1	0	0	0	1	0	0	Change in Long Term Debt to Long Term Debt
C9(3)	C9	0	0	0	0	1	1	1	0	0	1	Change in Total Debt to Total Debt
C10(3)	C10	0	1	1	0	1	0	0	1	1	0	Proceeds on Share Issue to Total Assets
C11(3)	C11	0	0	0	0	0	0	1	1	0	0	Cash Invested in Investing Activities to Fixed Assets
C12(3)	C12	0	0	0	1	0	0	0	0	0	0	Proportion of Dividend that is Non Cash
M1(3)	M1	0	0	0	0	0	0	0	1	0	0	Trading Turnover
M2(3)	M2		1	1	1	1	1	1	1	1	1	Total Dividend Yield
M3(3)	M3	1	1	1	0	1	1	1	0	1	0	Earnings Yield
M4(3)	M4	0	0	0	0	1	0	0	0	0	0	Market Value of Equity to Total Debt
M5(3)	M5	1	1	1	1	1	1	0	0	1	1	Price Book Ratio
M6(3)	M6	0	0	0	0	0	0	1	1	1	0	Capital to EBITDA Ratio
G1(3)	G1	0	1	1	1	1	0	0	1	0	0	Log(Assets)
G2(3)	G2	1	0	0	1	0	0	1	0	1	0	Log(Market Capitalisation)
G3(3)	G3	0	0	0	0	1	0	0	0	1	1	Retained Income to Total Assets Ratio
G4(3)	G4	1	1	1	0	1	0	0	1	1	0	Retention Ratio
G5(3)	G5	0	0	0	1	1	0	1	1	0	0	Growth Rate
G6(3)	G6	0	1	1	0	1	1	1	1	0	0	Capital Commitments to Fixed Assets Ratio
N1(3)	N1	0	1	1	1	1	0	1	0	0	1	Directors Shareholding
N2(3)	N2	0	0	0	1	1	0	1	0	1	0	Auditors Report Qualification
N3(3)	N3	0	0	0	1	0	1	0	1	0	0	Controlled by Another Company
N4	N4	0	0	0	1	0	0	0	1	1	0	Number of Years Listed
N5	N5	1	1	1	1	1	0	1	1	0	1	Number of Years in Existence
SI Values		100%	55%	65%	66%	75%	94%	77%	62%	80%	85%	

Appendix H.3. Binary masks for 3 Year Forward Forecasting Models

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	
L1(3)	L1	0	0	0	0	0	0	0	0	1	0	Current Ratio
L2(3)	L2	0	0	0	0	0	0	1	0	0	0	Quick Ratio
L3(3)	L3	1	0	0	0	0	0	1	1	0	0	Cash Ratio
L4(3)	L4	1	0	0	0	0	0	0	0	0	0	Non-Group Current Ratio
L5(3)	L5	1	0	0	0	0	0	0	1	0	0	Non Group Quick Ratio
L6(3)	L6	1	0	1	0	0	0	0	0	0	0	Cash to Total Assets Ratio
L7(3)	L7	0	0	0	0	0	0	0	0	0	0	Receivables Turnover
L8(3)	L8	0	0	0	0	0	0	0	0	0	0	Inventory Turnover
L9(3)	L9	0	1	1	1	0	0	1	1	1	0	Payables Turnover
L10(3)	L10	0	0	0	0	0	0	0	1	0	0	Cash Conversion Cycle
L11(3)	L11	1	0	0	0	0	0	1	0	1	0	Preference Dividend Declared?
E1(3)	E1	0	0	1	0	0	0	1	1	0	0	Total Asset Turnover
E2(3)	E2	0	0	0	0	0	0	0	0	0	0	Fixed Asset Turnover
E3(3)	E3	0	0	0	1	1	1	0	0	1	1	Equity Turnover
P1(3)	P1	1	0	1	1	1	1	1	1	1	1	Operating Profit Margin
P2(3)	P2	1	0	1	1	1	1	1	1	1	1	Net Profit Margin
P3(3)	P3	1	0	1	1	0	0	1	0	0	0	Return on Total Capital
P4(3)	P4	0	0	0	1	1	0	0	0	0	0	Return on Total Equity
P5(3)	P5	1	0	0	1	0	1	1	1	1	1	Return on Ordinary Equity
P6(3)	P6	1	1	1	0	0	0	0	0	0	0	Basic EPS
P7(3)	P7	0	1	0	1	1	1	1	0	0	1	Headline EPS
S1(3)	S1	1	1	1	1	1	1	1	0	0	1	Debt Equity Ratio
S2(3)	S2	0	0	0	0	0	0	0	0	1	0	Long Term Debt to Assets Ratio
S3(3)	S3	0	0	1	1	1	1	1	1	0	1	Interest Bearing Debt to Asset Ratio
S4(3)	S4	0	1	0	0	0	0	0	0	0	0	External Long Term Debt to Asset Ratio
S5(3)	S5	1	0	0	0	0	0	0	0	0	0	Financial Leverage
S6(3)	S6	1	0	0	0	0	0	0	0	0	0	Total Debt to Assets Ratio
S7(3)	S7	0	1	0	1	1	1	1	1	1	1	Total Debt and Contingencies to Assets Ratio
S8(3)	S8	0	0	0	1	0	0	0	0	0	0	Total Commitments to Assets Ratio
S9(3)	S9	0	0	0	0	0	0	0	0	0	0	External Debt to Assets Ratio
C1(3)	C1	1	0	1	0	0	0	0	0	0	1	Interest Coverage Ratio (inverse)
C2(3)	C2	0	1	0	0	0	0	0	1	1	0	Fixed Charge Coverage Ratio (inverse)
C3(3)	C3	0	1	1	0	0	0	0	1	0	0	Interest Expense to Cash Flow Ratio
C4(3)	C4	0	1	1	1	1	1	0	1	1	1	Cash Flow to Long Term Debt Ratio
C5(3)	C5	0	0	1	0	0	0	0	0	0	0	Cash Flow to Interest Bearing Debt (inverse)
C6(3)	C6	0	0	0	0	0	0	0	0	0	0	Cash Flow to Total Debt Ratio
C7(3)	C7	0	0	0	0	0	0	0	0	0	0	Cash Flow from Operations to Total Debt
C8(3)	C8	1	0	0	0	1	0	1	0	0	0	Change in Long Term Debt to Long Term Debt
C9(3)	C9	0	0	0	0	0	0	0	0	0	0	Change in Total Debt to Total Debt
C10(3)	C10	1	0	1	0	0	0	1	0	0	0	Proceeds on Share Issue to Total Assets
C11(3)	C11	0	0	0	0	0	0	0	0	0	0	Cash Invested in Investing Activities to Fixed Assets
C12(3)	C12	0	0	0	0	0	0	0	0	0	0	Proportion of Dividend that is Non-Cash
M1(3)	M1	1	0	1	0	0	0	1	0	0	0	Trading Turnover
M2(3)	M2	1	1	1	1	1	1	1	1	1	1	Total Dividend Yield
M3(3)	M3	0	0	0	0	0	0	0	0	0	0	Earnings Yield
M4(3)	M4	1	1	1	1	0	1	0	0	1	0	Market Value of Equity to Total Debt
M5(3)	M5	1	1	0	0	1	0	0	1	1	0	Price Book Ratio
M6(3)	M6	1	0	0	0	0	0	0	1	1	0	Capital to EBITDA Ratio
G1(3)	G1	0	0	0	0	0	0	0	0	0	0	Log(Assets)
G2(3)	G2	1	1	1	1	1	1	1	1	1	1	Log(Market Capitalisation)
G3(3)	G3	1	0	1	1	0	0	0	0	0	1	Retained Income to Total Assets Ratio
G4(3)	G4	1	0	1	0	1	1	0	1	0	1	Retention Ratio
G5(3)	G5	0	0	0	1	0	0	1	0	1	0	Growth Rate
G6(3)	G6	0	1	1	1	1	1	1	1	0	1	Capital Commitments to Fixed Assets Ratio
N1(3)	N1	0	0	1	0	0	0	0	0	0	0	Directors Shareholding
N2(3)	N2	0	0	1	1	1	0	0	0	0	1	Auditors Report Qualification
N3(3)	N3	1	0	1	1	0	1	0	0	0	1	Controlled by Another Company
N4	N4	0	0	0	0	0	0	0	0	0	0	Number of Years Listed
N5	N5	1	1	1	1	1	1	1	0	1	1	Number of Years in Existence
SI Values		82%	84%	84%	89%	90%	90%	85%	84%	84%	90%	

Appendix H.4. Optimal feature subset feature frequency analysis

	Model	1YrFwd			2Yr Fwd				3YrFwd			Tot
	Optimal Feature Subset	A	A	A	A	B	A	B	A	B	C	
	Year prior to failure	1	2	3	2	2	3	3	3	3	3	
L1	Current Ratio	0	1	1	0	0	0	0	0	0	0	2
L2	Quick Ratio	0	1	0	1	0	0	0	0	0	0	2
L3	Cash Ratio	1	0	0	0	1	1	0	0	0	0	3
L4	Non-Group Current Ratio	0	0	1	0	0	0	0	0	0	0	1
L5	Non-Group Quick Ratio	0	1	0	0	0	0	1	0	0	0	2
L6	Cash to Total Assets Ratio	1	0	1	0	1	0	0	0	0	0	3
L7	Receivables Turnover	1	1	0	0	1	0	0	0	0	0	3
L8	Inventory Turnover	1	1	0	0	1	0	0	0	0	0	3
L9	Payables Turnover	1	1	1	1	1	1	1	0	0	0	7
L10	Cash Conversion Cycle	1	0	0	1	0	0	0	0	0	0	2
L11	Preference Dividend Declared?	1	0	0	1	1	0	0	0	0	0	3
Tot		7	6	4	4	6	2	2	0	0	0	
E1	Total Asset Turnover	0	0	0	1	0	1	1	0	0	0	3
E2	Fixed Asset Turnover	0	1	0	1	1	0	1	0	0	0	4
E3	Equity Turnover	0	1	0	0	0	0	0	1	1	1	4
Tot		0	2	0	2	1	1	2	1	1	1	
P1	Operating Profit Margin	1	0	1	1	0	0	1	1	1	1	7
P2	Net Profit Margin	0	0	1	0	1	1	0	1	1	1	6
P3	Return on Total Capital	0	0	1	0	0	0	0	0	0	0	1
P4	Return on Total Equity	0	1	0	0	1	1	1	1	0	0	5
P5	Return on Ordinary Equity	1	0	0	1	1	1	0	0	1	1	6
P6	Basic EPS	0	0	1	1	1	0	1	0	0	0	4
P7	Headline EPS	0	0	1	0	1	0	0	1	1	1	5
Tot		2	1	5	3	5	3	3	4	4	4	
S1	Debt Equity Ratio	1	1	0	0	0	1	1	1	1	1	7
S2	Long Term Debt to Assets Ratio	0	1	0	1	0	1	1	0	0	0	4
S3	Interest Bearing Debt to Asset Ratio	0	0	0	1	0	1	1	1	1	1	6
S4	External Long Term Debt to Asset Ratio	1	0	0	0	1	0	1	0	0	0	3
S5	Financial Leverage	0	0	0	1	1	0	1	0	0	0	3
S6	Total Debt to Assets Ratio	1	0	1	0	1	0	0	0	0	0	3
S7	Total Debt and Contingencies to Assets Ratio	0	0	1	0	1	1	1	1	1	1	7
S8	Total Commitments to Assets Ratio	0	0	0	0	0	0	0	0	0	0	0
S9	External Debt to Assets Ratio	1	1	1	0	1	0	0	0	0	0	4
Tot		4	3	3	3	5	4	5	3	3	3	
C1	Interest Coverage Ratio (Inverse)	0	1	0	0	0	0	0	0	0	1	2
C2	Fixed Charge Coverage Ratio (Inverse)	0	1	0	0	1	0	0	0	0	0	2
C3	Interest Expense to Cash Flow Ratio	1	1	0	1	1	0	0	0	0	0	4
C4	Cash Flow to Long Term Debt Ratio	0	1	0	1	0	0	0	1	1	1	5
C5	Cash Flow to Interest Bearing Debt (Inverse)	0	0	0	0	0	0	0	0	0	0	0
C6	Cash Flow to Total Debt Ratio	0	0	1	0	0	0	0	0	0	0	1
C7	Cash Flow from Operations to Total Debt	1	1	1	1	1	1	1	0	0	0	7
C8	Change in Long Term Debt to Long Term Debt	1	0	1	1	0	1	0	1	0	0	5
C9	Change in Total Debt to Total Debt	0	0	0	1	0	0	1	0	0	0	2
C10	Proceeds on Share Issue to Total Assets	0	0	0	0	1	0	1	0	0	0	2
C11	Cash Invested in Investing Activities to Fixed Assets	0	0	0	0	1	0	0	0	0	0	1

C12	Proportion of Dividend that is Non-Cash	0	0	0	0	0	0	0	0	0	0	0
Tot		3	5	3	5	5	2	3	2	1	2	
M1	Trading Turnover	0	0	0	1	1	0	0	0	0	0	2
M2	Total Dividend Yield	1	1	1	0	0	1	1	1	1	1	8
M3	Earnings Yield	0	1	0	0	0	1	1	0	0	0	3
M4	Market Value of Equity to Total Debt	1	0	1	0	0	0	1	0	1	0	4
M5	Price Book Ratio	1	0	1	0	0	1	1	1	0	0	5
M6	Capital to EBITDA Ratio	0	1	1	1	1	0	0	0	0	0	4
M7	Excess Return (1 Fin Yr)	1	0		1	0	0	1				3
M8	Excess Return (2 Fin Yrs)											0
Tot		4	3	4	3	2	3	5	2	2	1	
G1	Log(Assets)	0	0	0	0	1	1	0	0	0	0	2
G2	Log(Market Capitalisation)	1	1	0	0	0	0	1	1	1	1	6
G3	Retained Income to Total Assets Ratio	1	0	0	0	0	1	1	0	0	1	4
G4	Retention Ratio	0	0	0	1	1	0	1	1	1	1	6
G5	Growth Rate	0	1	0	1	1	0	1	0	0	0	4
G6	Capital Commitments to Fixed Assets Ratio	0	1	0	0	0	0	1	1	1	1	5
Tot		2	3	0	2	3	2	5	3	3	4	
N1	Directors Shareholding	0	0	1	0	0	0	1	0	0	0	2
N2	Auditors Report Qualification	1	1	0	0	0	0	0	1	0	1	4
N3	Controlled by Another Company	0	0	1	0	1	0	0	0	1	1	4
N4	Number of Years Listed	1		1			1	1	0	0	0	4
N5	Number of Years in Existence			1					1	1	1	4
Tot		2	1	4	0	1	1	2	2	2	3	

APPENDIX I – MODEL PREDICTIONS

This appendix presents the predicted classes for each company as made by each model constructed in this study. Misclassifications are highlighted. Company references can be looked up in Appendix D.

		Forecast Period	1 Yr Fwd		2 Yr Fwd				3 Yr Fwd					
		Classifier	kNN	KRR	kNN		KRR		kNN			KRR		
Co. ref.	True Class	Optimal Feature Subset	A	A	A	B	A	B	A	B	C	A	B	C
N1	1		1	1	1	1	1	1	1	1	1	1	1	1
N2	1		1	1	1	1	1	1	1	1	1	1	1	1
N3	1		1	1	1	1	1	1	1	1	1	1	1	1
N5	1		[illegible]	[illegible]	1	1	1	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
N6	1		1	1	1	1	1	1	1	1	1	1	1	1
N8	1		1	1	1	1	1	1	1	1	1	1	1	1
N9	1		1	1	1	1	1	1	1	1	1	1	1	1
N10	1		1	1	1	1	1	1	1	1	1	1	1	1
N11	1		1	1	1	1	1	1	1	1	1	1	1	1
N14	1		1	1	1	1	1	1	1	1	1	1	1	1
N16	1		1	1	1	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
N17	1		[illegible]	[illegible]	1	[illegible]	1	[illegible]	1	1	1	1	1	1
N18	1		1	1	1	1	1	1	1	1	1	1	1	1
N19	1		1	1	1	1	1	1	1	1	1	1	1	1
N20	1		1	1	1	1	1	1	1	1	1	1	1	1
N21	1		1	1	1	1	1	1	1	1	1	1	1	1
N22	1		1	1	1	1	1	1	1	1	1	1	1	1
N23	1		1	1	1	1	1	1	1	1	1	1	1	1
N25	1		1	1	1	1	1	1	1	1	1	1	1	1
N26	1		1	1	1	1	1	1	1	1	1	1	1	1
N27	1		[illegible]	[illegible]	1	[illegible]	1	[illegible]	1	1	1	1	1	1
N28	1		1	1	1	1	1	1	1	1	1	1	1	1
N29	1		1	1	1	1	1	1	1	1	1	1	1	1
N30	1		1	1	1	1	1	1	1	1	1	1	1	1
N31	1		1	1	1	1	1	1	1	1	1	1	1	1
N32	1		1	1	1	1	1	1	1	1	1	1	1	1
N33	1		1	1	1	1	1	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
N34	1		1	1	1	1	1	1	1	1	1	1	1	1
N36	1		1	1	1	1	1	1	[illegible]	1	1	[illegible]	1	1
N38	1		1	1	[illegible]	1	[illegible]	1	1	1	1	1	1	1
N39	1		1	1	1	1	1	1	1	1	1	1	1	1
N41	1		1	1	1	1	1	1	1	1	1	1	1	1
N42	1		1	1	[illegible]	1	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
N43	1		1	1	1	[illegible]	1	[illegible]	1	1	1	1	1	1
N44	1		1	1	1	1	1	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
N46	1		1	1	1	1	1	1	1	1	1	1	1	1
N49	1		[illegible]	1	[illegible]	[illegible]	1	[illegible]	1	1	1	1	1	1
N50	1		1	1	1	1	1	1	1	1	1	1	1	1
N51	1		1	1	1	1	1	1	1	1	1	1	1	1
N52	1		1	1	1	1	1	1	1	1	1	1	1	1
N54	1		1	1	1	1	1	1	1	1	1	1	1	1
N55	1		1	1	1	1	1	[illegible]	1	1	1	1	1	1
N56	1		1	1	1	1	1	1	1	1	1	1	1	1
N57	1		1	1	1	1	1	1	1	1	1	1	1	1
N58	1		1	1	1	1	1	1	1	1	1	1	1	1
N59	1		1	1	1	1	1	1	1	1	1	1	1	1
N61	1		1	1	1	1	1	1	1	1	1	1	1	1
N63	1		1	1	1	[illegible]	1	[illegible]	1	1	1	1	1	1
N64	1		1	1	1	1	1	1	1	[illegible]	1	1	[illegible]	1
N66	1		1	1	1	1	1	1	1	1	1	1	1	1
N67	1		1	1	1	1	1	1	1	1	1	1	1	1
N68	1		1	1	1	1	1	1	1	1	1	1	1	1
N70	1		1	1	1	1	1	1	1	1	1	1	1	1
N71	1		1	1	1	1	1	1	1	1	1	1	1	1
N72	1		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	1	1	1	1
N73	1		1	1	1	1	1	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
N76	1		1	1	1	1	1	1	1	1	1	1	1	1
N77	1		[illegible]	[illegible]	1	1	1	1	1	1	1	1	1	1
N79	1		1	1	1	1	1	1	1	1	1	1	1	1
N80	1		1	1	1	1	1	1	1	1	1	1	1	1
N82	1		1	1	1	1	1	1	1	1	1	1	1	1

Co. ref	True Class	Forecast Period →	1 Yr Fwd		2 Yr Fwd				3 Yr Fwd					
		Classifier	kNN	KRR	kNN		KRR		kNN			KRR		
		Optimal Feature Subset	A	A	A	B	A	B	A	B	C	A	B	C
F1	-1		-1	-1	1	-1	1	-1	-1	-1	-1	-1	-1	-1
F2	-1		-1	-1	1	1	0	1	-1	-1	-1	-1	-1	-1
F3	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F5	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F6	-1		-1	-1	1	-1	1	-1	1	1	1	1	1	1
F8	-1		-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1
F9	-1		-1	-1	-1	1	-1	1	-1	-1	-1	-1	-1	-1
F10	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F11	-1		-1	-1	-1	1	1	1	1	1	1	1	1	1
F14	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F16	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F17	-1		-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1
F18	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F19	-1		1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F20	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F21	-1		-1	-1	-1	-1	-1	-1	1	1	1	1	1	1
F22	-1		1	1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1
F23	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	-1
F25	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F26	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F27	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F28	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F29	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F30	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F31	-1		1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
F32	-1		-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1
F33	-1		-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1
F34	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F35	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F36	-1		-1	-1	1	-1	1	-1	-1	-1	-1	-1	-1	-1
F39	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F41	-1		-1	-1	-1	-1	-1	-1	1	1	1	1	1	1
F42	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F43	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F44	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F46	-1		1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
F49	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F50	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F51	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F52	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F54	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F55	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F56	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F57	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F58	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F59	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F61	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F63	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F64	-1		-1	-1	1	-1	1	-1	-1	-1	-1	-1	-1	-1
F66	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F67	-1		-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1
F68	-1		1	1	1	-1	1	-1	-1	-1	-1	-1	-1	-1
F70	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F71	-1		-1	-1	-1	-1	-1	-1	1	1	-1	1	1	-1
F72	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F73	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F76	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F77	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F78	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F80	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
F82	-1		-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
		Total Err Rate	9.0%	8.2%	13.3%	12.3%	13.1%	13.9%	9.8%	9.8%	9.8%	10.7%	10.7%	9.8%
		Type I Err	9.8%	8.2%	6.6%	11.5%	4.9%	13.1%	11.5%	11.5%	9.8%	11.5%	11.5%	9.8%
		Type II Err	8.2%	8.2%	18.0%	13.1%	18.0%	14.8%	8.2%	8.2%	9.8%	9.8%	9.8%	9.8%